International Associ
of Geodesy Sym

Vo

F. Sans

A Window on the Futu

International Association of Geodesy Symposia

Fernando Sansò, Series Editor

International Association of Geodesy Symposia

Fernando Sansò, Series Editor

Symposium 101: Global and Regional Geodynamics
Symposium 102: Global Positioning System: An Overview
Symposium 103: Gravity, Gradiometry, and Gravimetry
Symposium 104: Sea SurfaceTopography and the Geoid
Symposium 105: Earth Rotation and Coordinate Reference Frames
Symposium 106: Determination of the Geoid: Present and Future
Symposium 107: Kinematic Systems in Geodesy, Surveying, and Remote Sensing
Symposium 108: Application of Geodesy to Engineering
Symposium 109: Permanent Satellite Tracking Networks for Geodesy and Geodynamics
Symposium 110: From Mars to Greenland: Charting Gravity with Space and Airborne Instruments
Symposium 111: Recent Geodetic and Gravimetric Research in Latin America
Symposium 112: Geodesy and Physics of the Earth: Geodetic Contributions to Geodynamics
Symposium 113: Gravity and Geoid
Symposium 114: Geodetic Theory Today
Symposium 115: GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications
Symposium 116: Global Gravity Field and Ist Temporal Variations
Symposium 117: Gravity, Geoid and Marine Geodesy
Symposium 118: Advances in Positioning and Reference Frames
Symposium 119: Geodesy on the Move
Symposium 120: Towards an Integrated Global Geodetic Observation System (IGGOS)
Symposium 121: Geodesy Beyond 2000: The Challenges of the First Decade
Symposium 122: IV Hotine-Marussi Symposium on Mathematical Geodesy
Symposium 123: Gravity, Geoid and Geodynamics 2000
Symposium 124: Vertical Reference Systems
Symposium 125: Vistas for Geodesy in the New Millennium
Symposium 126: Satellite Altimetry for Geodesy, Geophysics and Oceanography
Symposium 127: V Hotine Marussi Symposium on Mathematical Geodesy
Symposium 128: A Window on the Future of Geodesy

A Window on the Future of Geodesy

Proceedings of the International Association of Geodesy

IAG General Assembly
Sapporo, Japan
June 30 - July 11, 2003

Edited by
Fernando Sansò

Springer

Volume Editor

Prof. Dr. Fernando Sansò
Polytechnic of Milan
D.I.I.A.R. – Surveying Section
Piazza Leonardo da Vinci, 32
20133 Milan
Italy

Series Editor

Prof. Dr. Fernando Sansò
Polytechnic of Milan
D.I.I.A.R. – Surveying Section
Piazza Leonardo da Vinci, 32
20133 Milan
Italy

Library of Congress Control Number: 2004115723

ISSN 0939-9585
ISBN 3-540-24055-1 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springeronline.com
© Springer-Verlag Berlin Heidelberg 2005
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: design & production GmbH, Heidelberg
Production: Almas Schimmel
Typesetting: Camera ready by the authors
Printing: Mercedes-Druck, Berlin
Binding: Stein+Lehmann, Berlin

Printed on acid-free paper 32/3141/as 5 4 3 2 1 0

Preface

Every four years the International Association of Geodesy meets at the IUGG General Assembly and this has always been an important event for IAG to make the point on where are we going as geodesists both in terms of scientific production as well as in terms of organization.
The proceedings of IAG at the Sapporo 2003 General Assembly are the mirror of our scientific achievements, and, as Geodesy is a living entity like any other science, we could say it is a way to see the picture of what we consider our field of applications as well as of theoretical speculations. Let us examine this aspect in terms of what are: the object of our research, the methods we use, the general scientific results we can produce.

♦ **Our object:** here I would like to use a pseudo-Helmert definition; the object of Geodesy is knowing the surfaces of the earth: the geometric surface by positioning and e.m. surveying, and the physical surface, i.e the gravity field, by land, marine or satellite gravimetry, and their time variations.

This "object" is naturally interlaced with other physical properties of the earth both through deep processes affecting its surface and through the gravity field at all different scales from the global to the regional and local, where most engineering applications take place.

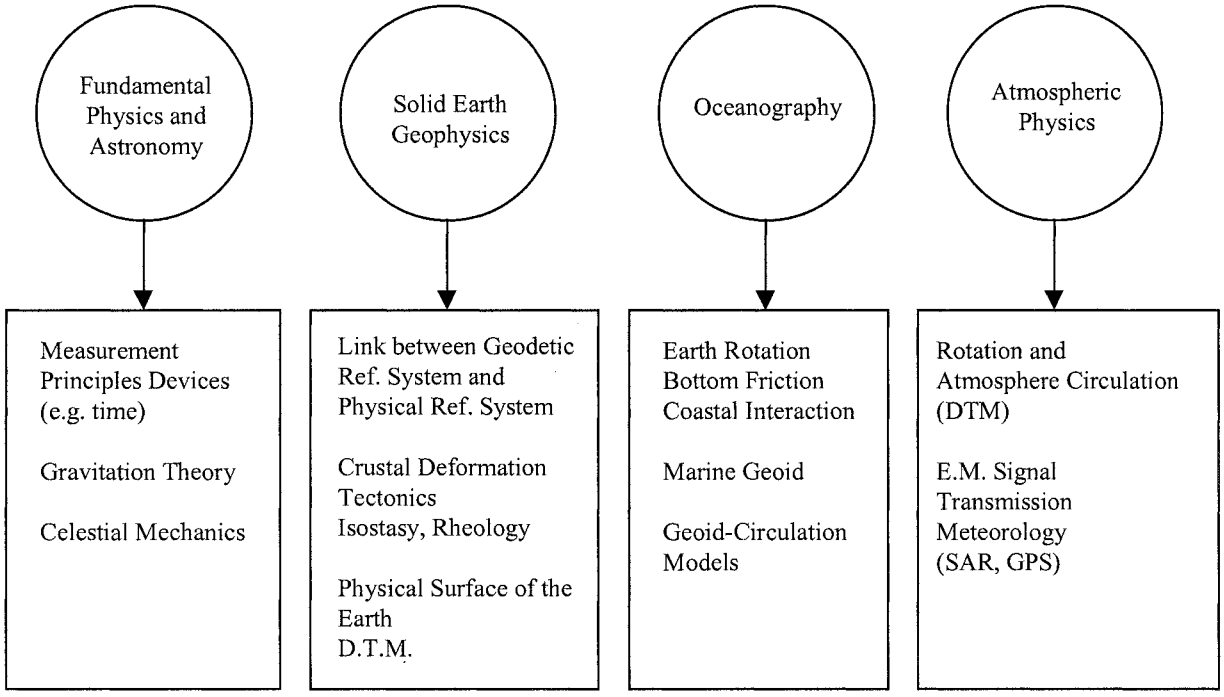
♦ **Our methods:** are typically those of modelling observation equations including as many physical effects as can be identified by specific measurement techniques and then studying the statistical nature of residuals by several means.

In the end we are then left with an optimal estimation problem where the unknowns can be either discrete in nature (sometimes even integers) or continuous like time signals or spatial fields, either governed by precise system equations or by approximated laws and henceforth either deterministic (at least in the average) or stochastic in full right.

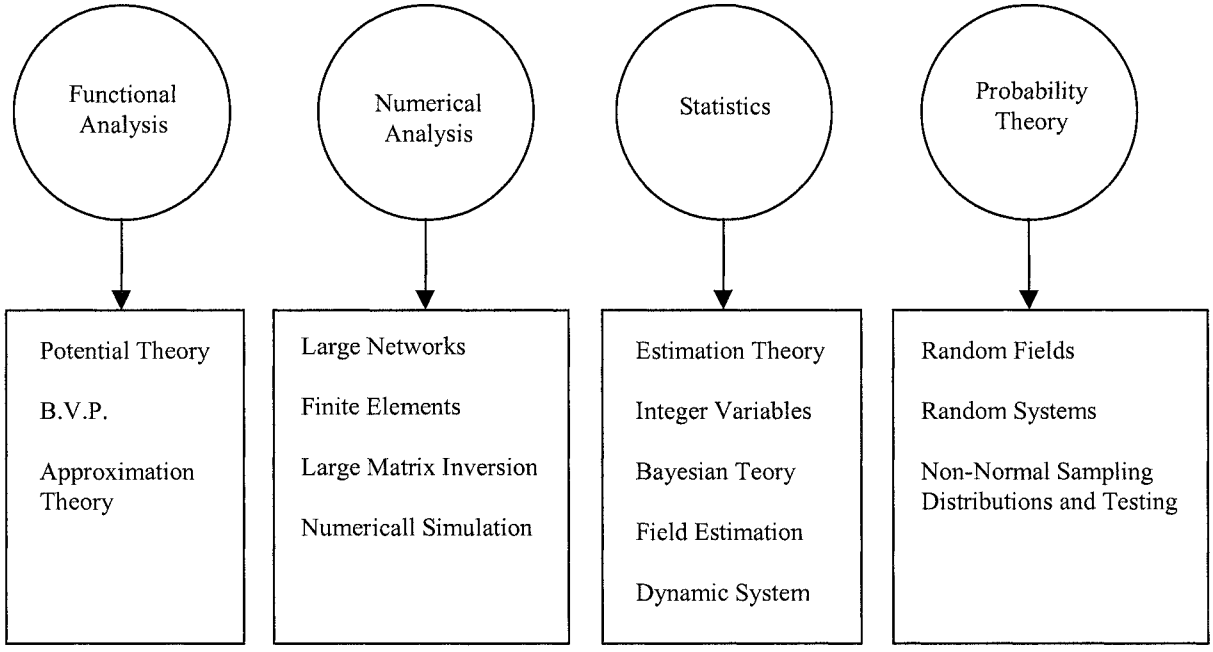
In manipulating an increasing catalogue of physical phenomena showing up into the observations of geodetic quantities we have enormously extended the baggage of knowledge which has to be in our arsenal of geodesists and the fields of their applications.

I have been trying to outline the main interfaces of Geodesy with other Sciences, which I have grouped into 3 families, Physics, Mathematics and Engineering. I'm sure this picture is far from being complete, however I hope it is representative.

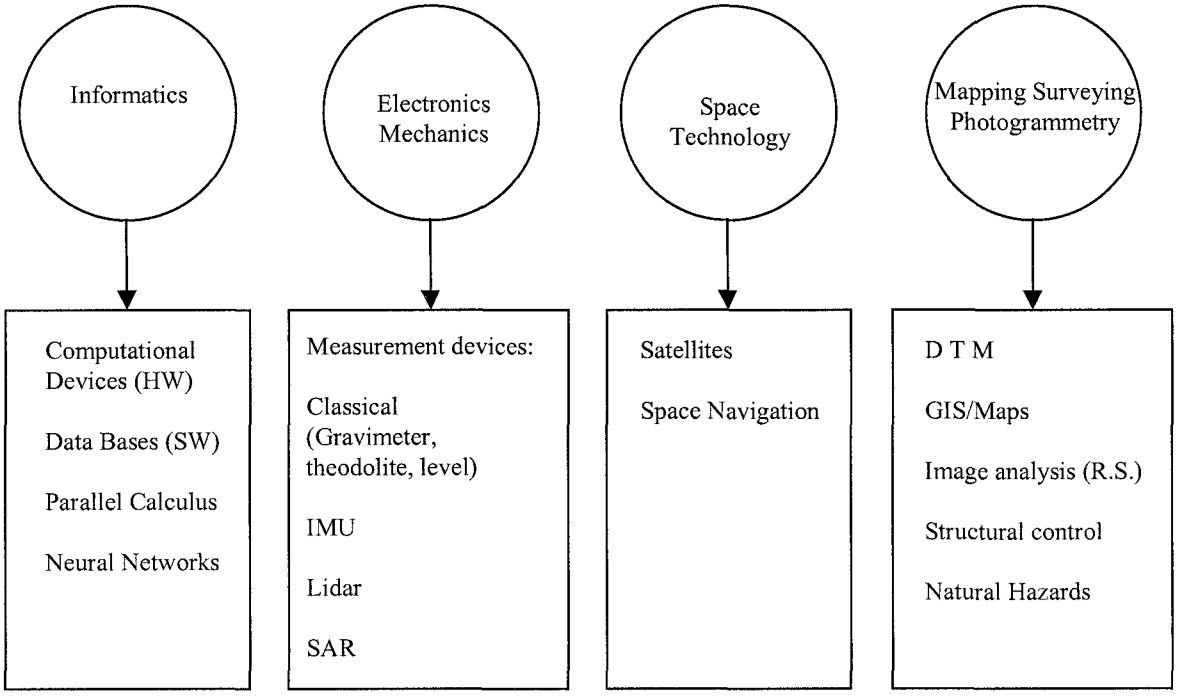
PHYSICS



MATHEMATICS



ENGINEERING



♦ **Our scientific products:** are indeed the knowledge of the objects of our investigation, i.e. the geometric surface of the earth and its gravity field jointly with their time variations.

In this respect we have to stress that knowing the geometric surface means also being able to attach precise coordinates to points lying on it and this can be done only by defining on the same time a unique reference system to which all positions have to be referred.
This is one of the most important general products of IAG, provided to all the other geo-sciences and techniques; namely how to refer any spatial event on the earth and in the surrounding space to a unique reference system attached to the earth.

In a very similar fashion, when we speak of the knowledge of the gravity field, particularly in terms of one of its equipotential surfaces, we implicitly mean that we are able to accomplish the unification of the many more or less local representations of it, for instance identifying in a unique mode the queen of these surfaces, the geoid, to serve as a unique reference surface for heights on the earth, i.e. the unified worldwide height datum.

Now that we have outlined what we think we are, we can analyze the directions from which we expect to receive strong impulses in the development of Geodesy, say in the next 10 years.

- Informatics, HW and SW development. In particular the exponential law for the development of computing power has not yet exhausted its trend and completely new technologies are close to enter into the market, opening the doors to more advanced calculations.

- More advanced technology in geodetic instrumentation with two main characteristics: one is the usual trend in improving accuracy, the other is to increase enormously the mass of information, also thanks to the diffusion of a large number of geodetic apparatus with may be a lower accuracy performance but also with a very low cost, which makes them accessible to a large community of users (as an example think of GPS).

- More advanced spatial technology for an improved surveying both at a global scale (think for instance of the new gravity missions CHAMP, GRACE and GOCE) and at a local scale (think for instance of the high resolution imagery or the InSAR surveying), to the effect of continuously growing the mass of information available to us; and all that, not only for the earth but also for the other planets, which are our job as well.

- More advanced knowledge of the physics of the solid earth, e.g. to facilitate the modelling of the crust deformation in geodetic surveying, of the ocean, e.g. in improving the knowledge of the global circulation pattern thus making more usable the satellite altimetry for the estimate of the marine geoid, of the atmosphere, e.g. providing better corrections for the effects of the propagation of our GPS signals,

- certainly a stronger progress in mathematics (e.g. in the area of dynamic systems or in that of random fields) as well as in statistics and data handling, driven by the great possibilities provided by the new electronic tools.

Of course I don't even attempt to be exhaustive but I think that the mentioned items will really impress a strong push to Geodesy. So we can ask now, what will Geodesy do for the other sciences?

First of all we shall pursue our general scopes and in that we will serve all the other geosciences by providing a global geodetic reference frame may be with an accuracy in the millimetric range and a unified height datum, may be in the centimetric range.

In this respect let me be clear; I know we are talking about the 1cm geoid since years, however we are still far from this goal in absolute sense and may be with the help of the new gravity satellite missions we will be able to approach that figure as an overall upper bound of the error.

Already being able to bridge at such a level of accuracy between the geometrically/kinematically defined reference frame and the physical reference systems, both in the sense of the gravity field and of the rotation of the earth, is an enormous step ahead in understanding the physical behaviour of this complicated "system earth" and Geodesy seems to be in pole position not only to provide the necessary experimental information but also to solve the difficult knot of modelling the interactions of the subsystems of the earth.

In addition we expect Geodesy to take the leadership of the new concept of continuously surveying the earth from space even at a regional level, exploiting its natural skill in combining different observation equations in a unique system.

To do this I'm sure that an improved structure of IAG Services will play a major role in providing data and specific products, for instance in such fields as engineering positioning and navigation, regional crustal deformation, digital elevation modelling, gravity and gravity variations surveying, atmospheric parameters monitoring, steady oceanic circulation and variability etc.

On the other hand Geodesy, with its new big challenge of optimal combination of different huge data sets, will be able to work out, as it happened in the past, original methods of field modelling and spatio/temporal signal analysis thus giving contributions to different areas of mathematics in terms of interesting problems and advanced solutions, for instance for items like satellite dynamics, boundary value problems, random fields analysis, general estimation theory with integer variables etc.

Remember that all the other sciences, including classical signal analysis, have been living for more than one century on the concept of least squares which has been worked out in a geodetic context and thus is in its full right a contribution of Geodesy to all other sciences.

I wish and I believe it will be possible that something similar will happen in the next years for instance by proposing a unified view on the very general item of field estimation, specially for inverse and improperly posed problems theory.

So for the moment it remains to me only to invite you to read the content of this book because I know that all these future developments are already starting now, and if you like to take part in them it is important that you know what is written here.

Fernando Sansò

Acknowledgement

The reviewing process of the scientific papers published in this proceedings volume has been performed by the conveners of the symposia; they have been acting as editors of individual chapters.
My warmest thanks should therefore be given to Chris Rizos, Alan H. Dodson, C.K. Shum, Pascal Willis, Michael G. Sideris, Bernhard Heck and Clark Wilson.

A sincere thanks are also due to the Local Organizing Committee, that has made of the General Assembly such a thoughtful and enjoyable event.

Contents

Symposium G01: Positioning 1
Conveners: C. Rizos, A.H. Dodson

ITRF Combination – Status and Recommendations for the Future 3
D. Angermann, H. Drewes, M. Gerstl, R. Kelm, M. Krügel, B. Meisel

Time Evolution of the Terrestrial Reference Frame 9
D. Angermann, B. Meisel, M. Krügel, H. Müller, V. Tesmer

Scale Variation of GPS Time Series 15
M. Poutanen, J. Jokela, M. Ollikainen, H. Koivula, M. Bilker, H. Virtanen

Vertical Repeatability of GPS Measurements in the Tsukuba GPS/MET Dense-Network Campaigns 21
S. Shimada

Horizontal Crustal Motion Estimates for the UK Inferred from GPS: Preliminary Results 26
A.I. Waugh, F.N. Teferle, R.M. Bingley, A.H. Dodson

Results of the SIRGAS Campaign 2000 and Coordinates Variations with Respect to the 1995 South American Geocentric Reference Frame 32
H. Drewes, K. Kaniuth, C. Völksen, S.M. Alves Costa, L.P. Souto Fortes

Implementation of a Semi-Dynamic Datum for New Zealand 38
G. Blick, C. Crook, D. Grant, J. Beavan

GPS Measurements to Investigate the Reason why GPS is Less Accurate in Mountain Areas 44
M. Satomura, S. Shimada, Y. Goto, M. Nishikori

Towards More Reliable Estimation of GPS Positioning Accuracy 48
J. Krynski, Y.M. Zanimonskiy

Realistic Uncertainty Measures for GPS Observations 54
S. Schön, H. Kutterer

Networked Transport of RTCM via Internet Protocol (NTRIP) 60
G. Weber, D. Dettmering, H. Gebhard

Numerical Weather Predictions for GPS Positioning 65
A.B.O. Jensen

A Multi-Base-Station Approach for Long Range Differential GNSS Positioning of Airborne Sensors 71
N.S. Kjorsvik, O. Ovstedal, J.G.G. Svendsen, L.E. Blankenberg

Point Real-Time Kinematic Positioning 77
Y. Gao, M. Abdel-Salam, K. Chen, A. Wojciechowski

The Integration of GPS and Pseudolites for Bridge Monitoring 83
J. Barnes, C. Rizos, H-K. Lee, G.W. Roberts, X. Meng, E. Cosser, A.H. Dodson

Hybrid Sensor System for Bridge Deformation Monitoring: Interfacing with Structural Engineers 89
X. Meng, A. Dodson, G. Roberts, E. Cosser

A Fuzzy System for the Assessment of Landslide Monitoring Data 95
M. Haberler

OASYS – Integrated Optimization of Landslide Alert Systems 101
H. Kahmen, W. Niemeier

Combined Evaluation of Geodetic and Geotechnical Data during Tunnel Excavation by Use of a Knowledge-Based System 105
K. Chmelina, H. Kahmen

Symposium G02: Advanced Space Technology 111
Conveners: C.K. Shum, P. Willis

Quasi Real-Time Positioning of Spacecrafts Using the Internet VLBI System 113
T. Kondo, Y. Koyama, R. Ichikawa, M. Sekido, H. Osaki

Rapid Turnaround EOP Measurements by VLBI over the Internet 119
Y. Koyama, T. Kondo, H. Osaki, A.R. Whitney, K.A. Dudevoir

Determination of Tropospheric Parameters Within the New IVS Pilot Project 125
H. Schuh, J. Boehm, G. Engelhardt, D. MacMillan, R. Lanotte, P. Tomasi, M. Negusini, I. Vereshchagina, V. Gubanov, R. Haas

VLBIONOS - Probing the Ionosphere by Means of Very Long Baseline Interferometry 131
T. Hobiger, J. Boehm, H. Schuh

Numerical Simulations for Geodetic Parameter Estimation from Space VLBI Data 137
M.N. Kulkarni, P. Pillai, S. Frey

Relativistic VLBI Delay Model for Finite Distance Radio Source 141
M. Sekido, T. Fukushima

Synergy of VLBI and GPS Observations in Japan 146
H. Tsuji, T. Tanabe, H. Kawawa, K. Miyagawa, K. Takashima, S. Kurihara, Y. Fukuzaki, S. Matsuzaka

Ground-Based Verification of Four-Way Doppler Measurement for the SELENE Mission 152
H. Noda, K. Asari, N. Namiki, T. Iwata, Y. Kono, H. Hanada, N. Kawano, Z. Yamamoto

Results of the Critical Design of RSAT/VRAD Mission Instruments on SELENE Sub-Satellites Rstar/Vstar for Selenodesy 157
T. Iwata, T. Sasaki, T. Izumi, Y. Kono, H. Hanada, N. Kawano, F. Kikuchi

Application of a PZT Telescope to In situ Lunar Orientation Measurement (ILOM) 163
H. Hanada, K. Heki, H. Araki, K. Matsumoto, H. Noda, N. Kawano, T. Tsubokawa, S. Tsuruta, S. Tazawa, K. Asari, Y. Kono, T. Yano, N. Gouda, T. Iwata, T. Yokoyama, H. Kanamori, K. Funazaki, T. Miyazaki

Combined Processing of GPS Data from Ground Station and LEO Receivers in a Global Solution 169
U. Hugentobler, A. Jäggi, S. Schaer, G. Beutler

Efficient Stochastic Orbit Modeling Techniques using Least Squares Estimators 175
A. Jäggi, G. Beutler, U. Hugentobler

Kinematic Precise Orbit Determination for Gravity Field Determination 181
D. Švehla, M. Rothacher

Gravity Field Recovery and Validation by Analysis of Short Arcs of a Satellite-to-Satellite Tracking Experiment as CHAMP and GRACE 189
K.H. Ilk, M. Feuchtinger, T. Mayer-Gürr

Combination of Sea-Level Observations and an Oceanographic Model for Geodetic Applications in the Baltic Sea 195
K. Novotny, G. Liebsch, R. Dietrich

Analysis of GPS Campaign Observations of the Asia Pacific Regional Geodetic Project (APRGP) - Establishment of a Velocity Field - 201
Y. Tanaka, T. Amagai, A. Yamada, S. Matsuzaka, H. Tsuji

The Proposed International DORIS Service 207
P. Willis, G. Tavernier, M. Feissel-Vernier, F. Lemoine, C. Noll, J. Ries, L. Soudarin

Calibration/Validation of GOCE Data by Terrestrial Torsion Balance Observations 214
G. Tóth, J. Ádám, L. Földváry, I.N. Tziavos, H. Denker

Symposium G03: Determination of the Gravity Field 221
Convener: M.G. Sideris

Gravimetry and Gravity Instrumentation

Repeated Measurements of Gravity with the Absolute Gravimeter FG5 #210 at Matsushiro, Japan and Comparison with the Superconducting Gravimeter T011 223
Y. Imanishi, T. Higashi, Y. Fukuda, K. Kokubo

Application of Gravimetry by Helicopter to Identify Marine Active Faults and Improve Accuracy of Geoid at Coastal Zones 229
J. Segawa, J.E. Joseph, E. Nakayama, Kumar V.K., S. Kusumoto, T. Ito, S. Sekizaki, T. Ishihara, M. Komazawa

A Gravity Database of Southwest Japan: Application to Bouguer Gravity Imaging in Kyushu District, Southwest Japan 236
R. Shichi, A. Yamamoto, T. Kudo, Y. Murata, K. Nawa, M. Komazawa, M. Nakada, H. Miyamachi, H. Komuro, Y. Fukuda, T. Higashi, Y. Yusa, I. Nakagawa, H. Watanabe, J. Oikawa, S. Kobayashi, I. Ohno

Gravity-Based Fault Mapping: the Ishikari Lowland, Hokkaido, Japan 242
A. Yamamoto

Compilation and Evaluation of a Consistent Marine Gravity Data Set Surrounding Europe 248
H. Denker, M. Roland

Comparison of Downward Continuation Methods of Airborne Gravimetry Data 254
F. Mueller, T. Mayer-Guerr

Optimisation of Density Parameter in Gravity Prediction from Bathymetry 259
D.M. Sproule, A.H.W. Kearsley

Gravity Satellite Missions: Data processing and applications

Calibration and Validation of GOCE Gravity Gradients 265
J. Bouman, R. Koop, R. Haagmans, J. Müller, N. Sneeuw, C.C. Tscherning, P. Visser

Impact of Space-Borne Gravity Gradiometer Instrument Filter on Observation Error and Gravity Field Recovery Performance for GOCE 271
P.N.A.M Visser

Testing Frame-Transformation, Gridding and Filtering of GOCE Gradiometer Data by Least-Squares Collocation Using Simulated Data 277
C.C. Tscherning

Processing of CHAMP ACC and RSO Data in the Earth Gravity Field Determination 283
J. Luo, J. Sh. Ning

One Year of Time-Variable CHAMP-Only Gravity Field Models Using Kinematic Orbits 288
N. Sneeuw, C. Gerlach, L. Földváry, T. Gruber, T. Peters, R. Rummel, D. Švehla

Topographic-Isostatic Reductions in Satellite Gravity Gradiometry Based on a Generalized Condensation Model 294
B. Heck, F. Wild

Long-Wavelength Discrepancies Between the Australian Gravity Field and EIGEN-2 CHAMP Data 300
W.E. Featherstone

Comparison of new Geoid Models and EIGEN-2S in the North Atlantic Region 306
O.C.D. Omang, R. Forsberg, G. Strykowski

Computation of a Geopotential Model from GOCE Data Using Fast Spherical Collocation - A Simulation Study 310
C.C. Tscherning, D.N. Arabelos

Satellite Altimetry for Gravity, Geoid and SST

Improved Determination of Sea Surface Heights Close to the Australian Coast from ERS-2 Satellite Radar Altimetry 314
X. Deng, W.E. Featherstone

A Numerical Study of Solving the Altimetry-Gravimetry Boundary Value Problem in Coastal Regions 320
R.S. Grebenitcharsky, M.G. Sideris

Improved High Resolution Altimetric Gravity Field Mapping (KMS2002 Global Marine Gravity Field) 326
O.B. Andersen, P. Knudsen, R. Trimmer

On the Determination of Marine Geoid Models by Least-Squares Collocation and Spectral Methods Using Heterogeneous Data 332
G.S. Vergos, I.N. Tziavos, V.D. Andritsanos

Heights and Vertical Datums

Progress Towards Implementation and Development of a New Zealand National Vertical Datum 338
M.J. Amos, W.E. Featherstone, G.H. Blick

Prediction of Vertical Gravity Gradients Using Gravity and Elevation Data 344
Sz. Rózsa, Gy. Tóth

Geoid Modelling

Construction of a Synthetic Earth Gravity Model by Forward Gravity Modelling 350
M. Kuhn, W.E. Featherstone

First Results Towards an Isostatically Compensated Reference Earth Model 356
D. Tsoulis, B. Stary

Towards the Estimation of a Multi-Resolution Representation of the Gravity Field Based on Spherical Wavelets 362
M. Schmidt, O. Fabert, C.K. Shum

Stokes Integration versus Wavelet Techniques for Regional Geoid Modelling 368
M. Roland, H. Denker

A Preliminary Geoid Model for Africa 374
C.L. Merry, D. Blitzkow, H. Abd-Elmotaal, H.H. Fashir, S. John, F. Podmore, J.D. Fairhead

Optimal Marine Geoid Determination in the Atlantic Coastal Region of Argentina 380
C. Tocho, G.S. Vergos, M.G. Sideris

Comparison of Newton's Integral in the Space and Frequency Domains 386
M. Kuhn, K. Seitz

Wavelet Evaluation of Some Singular Geodetic Integrals 392
Q. Liu, M.G. Sideris

Gravimetric Geoid Determination in the Andes 398
C. Tocho, G. Font, M.G. Sideris

On the Numerical Solution of Parameters of the Least Squares Modification of Stokes' Formula 403
A. Ellmann

A Numerical Comparison of Different Ellipsoidal Corrections to Stokes' Formula 409
A. Ellmann

Symposium G04: General Theory and Methodology 415
Convener: B. Heck

On Total Least-Squares Adjustment with Constraints 417
B. Schaffrin, Y.A. Felus

GNSS Best Integer Equivariant Estimation 422
P.J.G. Teunissen

Performance Comparison of the BIE Estimator with the Float and Fixed GNSS Ambiguity Estimators 428
S. Verhagen, P.J.G. Teunissen

Parametric Versus Non-Parametric Methods for Optimal Weighted Averaging of Noisy Data Sets 434
C. Kotsakis, I.N. Tziavos

Estimation of Variance Components Through a Combined Adjustment of GPS, Geoid and Levelling Data 440
G. Fotopoulos, C. Kotsakis, M.G. Sideris

A Minimum Energy Condition for the Inverse Gravimetric Problem 446
O. Jamet, M. Diament

Successive Approximations in the Solution of a Weakly Formulated Geodetic Boundary-Value Problem 452
P. Holota

Computation of Geopotential Coefficients from Gravity Anomalies on the Ellipsoid 459
S.J. Claessens, W.E. Featherstone

Advanced Harmonic Development of the Earth Tide Generating Potential 465
S.M. Kudryavtsev

Global Isotropic Isostatic Response Function in Terms of Zonal Spherical Harmonics 471
H.A. Abd-Elmotaal

Paraná Basin: One Istostatic Modeling Using a Geographic Information System to Delineate the Main Basement Structures 477
M.C.L. Quintas

Patching Local Empirical Covariance Functions – A Problem in Altimeter Data Processing 483
P. Knudsen

Asymptotic Dislocation Theory of a Spherical Earth for Calculating Co-Seismic Displacement 488
W. Sun, S. Okubo

The Use of Energy Balance Relations for Validation of Gravity Field Models and Orbit Determination Results 494
K.H. Ilk, A. Löcher

Biharmonic Spline Wavelets Versus Generalized Multi-Quadrics for Continuous Surface Representations 500
R. Mautz, B. Schaffrin, J. Kaschenz

An Energy Approach for Intelligent Structural Health Monitoring Using Wavelet Analysis 506
N. El-Sheimy, M.M. Reda Taha, A. Osman, A. Noureldin

On Discrete Schemes in Downward Continuation of Gravity 512
W. Sun

Application of Clebsch-Gordan Coefficients and Isomorphic Frame Transformations to Invert Earth's Changing Geometrical Shape for Continental Hydrological Loading and Sea Level's Passive Response 518
G. Blewitt, P. Clarke, D. Lavallée, K. Nurutdinov

Role of Algebra in Modern Day Geodesy 524
J.L. Awange, Y. Fukuda, S. Takemoto, E.W. Grafarend

Symposium G05: Geodynamics 531
Convener: C.R. Wilson

GPS Survey theory and design

Mixed-Mode GPS Deformation Monitoring – A Cost-effective and Accurate Alternative? 533
V. Janssen, C. Rizos

Two Examples of Optimal Design of Geodynamic GPS Networks 538
N.V. Shestakov, H.E. Waithaka, M. Kasahara, M.D. Gerasimenko

Applications of GPS to Crustal deformation

Deformation of the South American Crust Estimated from Finite Element and Collocation Methods 544
H. Drewes, O. Heidbach

Present-Day Movement and Tectonic Strain Rate Field of Active Crustal Blocks in China Continent 550
C. Xu, H. Wang, J. Xu

Crustal Deformations at Permanent GPS Sites in Denmark 556
S.A. Khan, P. Knudsen, C.C. Tscherning

Monitoring Land Subsidence of Jakarta (Indonesia) Using Leveling, GPS Survey and InSAR Techniques 561
H.Z. Abidin, H. Andreas, M. Gamal, R. Djaja, C. Subarya, K. Hirose, Y. Maruyama, D. Murdohardono, H. Rajiyowiryono

Gravity applications

Microgravity Survey in the Southern Tip of the Atera Fault, Central Japan 567
T. Tanaka, H. Aoki, K. Oshita, K. Nozaki

Geothermal Reservoir Monitoring with Absolute Gravimetry 571
M. Sugihara

Earth Rotation and Geodynamics Techniques

Towards a Rigorous Combination of VLBI and GPS Using the CONT02 Campaign 576
D. Thaller, R. Schmid, M. Rothacher, V. Tesmer, D. Angermann

Fuzzy Inference Systems for the Prediction of Earth Rotation Parameters 582
O. Akyilmaz, H. Kutterer

Symposium G06: Insight into Earth System Science: Variations in the Earth's Rotation and its Gravity Field
Convener: C.R. Wilson 589

Earth rotation interpretation

Oceanic Excitation of Polar Motion from Intraseasonal to Decadal Periods 591
A. Brzeziński, J. Nastula, B. Kolaczek, R.M. Ponte

Excitations of Polar Motion from an Ensemble of Global Atmospheric Models 597
J. Nastula, D. Salstein, B. Kolaczek

Theoretical studies related to earth rotation and time variable gravity

Sensitivity Analysis of the Non-Linear Liouville Equation 601
F. Seitz, H. Kutterer

Filters to Estimate Water Storage Variations from GRACE 607
K.-W. Seo, C.R. Wilson, J. Chen, J.S. Famiglietti, M. Rodell

Author Index 613

Keyword Index 617

SYMPOSIUM G01

Positioning

C. Rizos
A.H. Dodson

ITRF Combination - Status and Recommendations for the Future

D. Angermann, H. Drewes, M. Gerstl, R. Kelm, M. Krügel, B. Meisel
Deutsches Geodätisches Forschungsinstitut (DGFI), Marstallplatz 8, D-80539 München, Germany
email : angerman@dgfi.badw.de

Abstract. In its function as ITRS Combination Center and IERS Combination Research Center, DGFI is involved in the combination of space geodetic observations. Based on the latest space geodetic solutions provided by individual analysis centers and/or technique centers, DGFI has computed a combined solution for station positions and velocities. Major goals are: (1) a validation of the various components of the ITRS Combination Center; (2) a verification and enhancement of the combination strategy; (3) an external comparison with ITRF2000; (4) the identification of still existing deficiencies; and (5) proposals for future ITRF realisations.

Keywords. ITRF, combination strategies, deficiencies, recommendations

1 Introduction

Since 1988, the International Earth Rotation and Reference Systems Service (IERS) has been in charge of the establishment and maintenance of the International Terrestrial Reference Frame (ITRF), a realisation of the International Terrestrial Reference System (ITRS). Conventionally, the ITRF is realised by the positions and constant velocities for a set of global tracking sites. The positions and velocities of these ITRF stations are derived from a combination of individual technique solutions provided by various analysis centers. The space geodetic techniques used at present are Very Long Baseline Interferometry (VLBI), Satellite and Lunar Laser Ranging (SLR/LLR), the Global Positioning System (GPS) and the Doppler Orbitography and Radio positioning Integrated by Satellite (DORIS).

A series of ten ITRFs were compiled by the responsible ITRS Product Center (Institute Géographique National, IGN, Paris), from ITRF88 to ITRF2000 (e.g. Boucher et al., 1999). The latest is the most extensive and accurate terrestrial reference frame, consisting of the positions and velocities of about 800 stations located at approximately 500 sites, with better distribution compared to past ITRF realisations (Altamimi et al., 2002).

The ITRF solutions are widely used in geodetic and geophysical applications and research projects. A highly accurate and consistent realisation of the terrestrial reference system is a fundamental prerequisite for progress in Earth sciences, and to support a multitude of new satellite missions (e.g. GRACE, JASON-1, ENVISAT, as well as for future GOCE and GALILEO) to achieve their scientific goals.

Within the newly re-organised IERS structure, the ITRS Product Center (IGN, Paris) is supplemented by ITRS Combination Centers, which were included as new IERS components to ensure redundancy for the ITRF computations. Currently there are three ITRS Combination Centers at DGFI, IGN, and NR-Can (National Resources Canada).

2 ITRS Combination Center at DGFI

The data flow and major components of the ITRS Combination Center at DGFI is shown in Figure 1. The combination methodology is based on the following:

- Validation and analysis of individual solutions: The available solutions are checked concerning various aspects, such as format and their suitability for a consistent TRF combination.

- Reconstruction of unconstrained normal equations: For the combination strategy on the level of free normal equations, we have to remove the a-priori datum constraints which are normally applied in the solutions.

- Preprocessing: For some of the input solutions it may be necessary to perform preprocessing, e.g. epoch transformations for station positions, renaming and reducing parameters.

- Datum definition: For the intra- and inter-technique combination, as well as for the final TRF computation, the datum definition is a key issue. In order to generate comparable solutions we define the datum in a consistent way by applying minimum constraints using no-net rotation and/or no-net translation conditions (depending on the degrees of freedom of the particular space technique). For this purpose we use a subset of "good" and well distributed stations for each technique (so-called reference stations).

- Intra-technique comparison and combination: Input are individual (unconstrained) normal equations of the same observation technique. Before combining them, a careful comparison is essential to identifiy possible problems, which could cause systematic effects (biases) in the combined results. Other important tasks are the identification and rejection of outliers, and the computation of relative weight factors for the individual normal equations.

- Inter-technique comparison and combination: Input are the combined (unconstrained) intra-technique normal equations resulting from the previous processing step. Key elements are co-location sites and local ties, to realise a unique frame, and to compare the solutions of the different techniques. Major tasks include the estimation of weight factors (e.g. by means of variance component estimation) for the normal equations of the different techniques, handling of local ties, equating velocities at co-location sites, datum definition, etc.

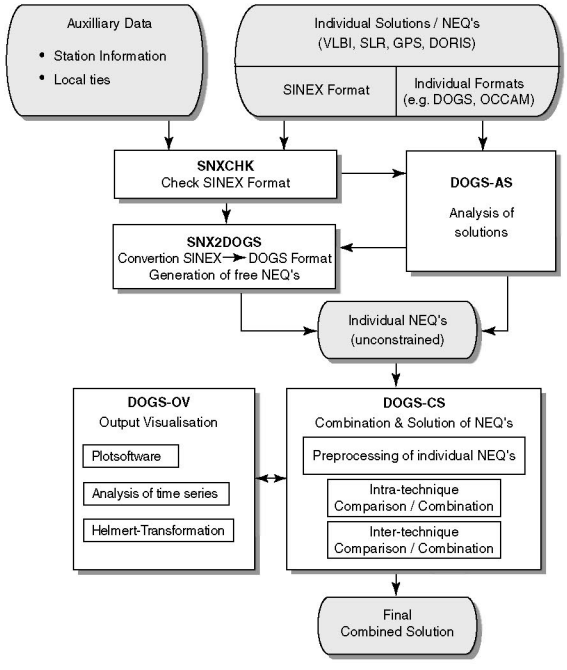


Fig. 1: Data flow and combination procedure.

3 Combined TRF solution and comparison with ITRF2000

The combined TRF solution was computed from multi-years solutions with station positions and velocities, submitted by several analysis centers for the ITRF2000 computation. In addition we used some newer solutions containing more recent observations (see Table 1). We applied the combination strategy described in the previous section.

Table 1: Summary of solutions used for combination. We excluded stations with observation spans less than 1 year. The numbers in parantheses indicate: (1) unconstrained normal equations; (2) loose constrained solutions, (3) minimum constrained solutions.

AC/Solution	Data Span	# Stations orig/used	Source
DORIS			
(GRGS)00D01	1993–1998	70/69	ITRF2000 (3)
(IGN)02D04	1993–2002	111/109	IGN/CDDIS (2)
GPS			
(IGS)03P01	1996–2002	216/207	NRCan (3)
SLR			
(CRL)00L02	1990–2000	62/62	ITRF2000 (2)
(CSR)00L04	1976–2000	141/106	ITRF2000 (2)
(DGFI)01L01	1981–2001	113/96	DGFI (1)
(JCET)00L05	1993–2000	63/55	ITRF2000 (2)
VLBI			
(DGFI)02R02	1984–2002	49/49	DGFI (1)
(GIUB)00R01	1984–1999	53/53	ITRF2000 (2)
(GSFC)00R01	1979–1999	138/88	ITRF2000 (2)
(SHA)00R01	1979–1999	129/88	ITRF2000 (2)

3.1 Intra-technique combination

The combination of data (solutions) of the same observation technique should be done by the corresponding technique centers (IGS, ILRS, IVS, IDS). At present, combined multi-years solutions with station positions and velocities are only produced by the IGS. We used the cummulative IGS solution (IGS03P01) provided by Natural Resources Canada

(NRCan, see Ferland, 2002), instead of individual GPS solutions.

In the case of SLR, we used four solutions provided by the CRL (Communications Research Laboratory, Japan), CSR (Center for Space Research, USA), JCET (Joint Center for Earth System Technology, USA) and DGFI. As an example the Helmert transformation results between the individual and combined SLR solutions are shown in Table 2. The results demonstrate the high stability of SLR with respect to the realisation of TRF origin and scale.

The four VLBI solutions provided by GIUB (Geodetic Institute of University Bonn, Germany), GSFC (Goddard Space Flight Center, USA), SHA (Shanghai Astronomical Observatory, China) and DGFI demonstrate the high quality of VLBI for the realisation of the TRF scale and to estimate precise station positions and velocities (see Table 3).

The DORIS solutions of GRGS (Groupe de Recherches de Géodésie Spatiale, France) and IGN (Institute Géographique National, France) agree to about 7-8 mm and 2 mm/yr for station positions and velocities, respectively. Unfortunately, the a-priori datum conditions could not be removed from the DORIS solutions.

Table 2: Helmert transformation results of individual SLR solutions w.r.t. combined intra-technique solution.

H.-T. Results	CRL	CSR	DGFI	JCET
Tx [mm]	-0.1	1.4	0.9	2.4
Ty [mm]	1.5	-0.6	-1.3	-5.0
Tz [mm]	0.3	1.0	0.2	0.8
Scale [ppb]	0.03	0.09	-0.26	-0.10
Tx rate [mm/yr]	0.8	0.1	-0.2	-0.4
Ty rate [mm/yr]	-1.4	-0.1	0.1	1.1
Tz rate [mm/yr]	-1.6	-0.3	1.1	-0.1
Scale rate [ppb/yr]	-0.09	0.04	-0.08	0.18
Pos RMS [mm]	4.1	4.9	2.3	3.7
Vel RMS [mm/yr]	1.6	1.4	0.7	1.2

Table 3: Helmert transformation results of individual VLBI solutions w.r.t. combined intra-technique solution.

H.-T. Results	DGFI	GIUB	GSFC	SHA
Scale [ppb]	0.17	-0.15	-0.34	-0.06
Scale rate [ppb/yr]	0.03	-0.08	-0.01	-0.02
Pos RMS [mm]	3.4	2.7	2.2	3.1
Vel RMS [mm/yr]	0.43	0.72	0.22	0.81

3.2 Inter-technique combination

In the inter-technique combination co-location sites (see Figure 2) and intra-site vectors (local ties) play a dominant role. Are the local ties sufficiently well determined to introduce them as constraints, and are different velocity estimations at the same site identical? To find the answer, it is essential to validate the local ties and the velocity estimations at co-location sites, before combining different techniques.

Table 4 summarises some of these validation results. A comparison between the local ties and computed coordinate differences obtained from the intra-technique solutions reveals for six co-locations (3 GPS-VLBI, 3 GPS-SLR) an excellent agreement (the absolute differences are below 5 mm). But on the other hand, there are many other co-locations (especially those with DORIS), where the differences exceed 2 cm. The situation is similar regarding station velocities between different techniques at co-location sites. For eight co-locations the absolute velocity differences are below 1mm/yr, whereas in many other cases the differences exceed 5 mm/yr.

Any interpretation of these discrepancies is difficult since various factors need to be considered (e.g. systematic biases between space geodetic solutions, non-linear site motions, local site-dependent effects, errors in local tie measurements, small inconsistencies related to the datum definition). Based on these comparisons we selected suitable local ties and we identified sites suitable for equating velocities. Then we introduced pseudo-observations with appropriate weights. To generate the final combined TRF solution we realised the geodetic datum by applying no-net rotation conditions with respect to ITRF2000 station positions and velocities. The origin (translations and their rates) is realised by SLR, and the scale and its rate by SLR and VLBI.

3.3 Comparison with ITRF2000

We compared the combined DGFI solution with ITRF2000 by means of a 14 parameter Helmert transformation. The results are summarised in Table 5. Both TRF solutions agree quite well in origin and scale, but it has to be considered that the datum of both TRF realisations was defined in a similar way, thus this is not a completely independent outcome. The RMS residuals of the reference stations for positions and velocities are displayed for each technique separately. Figure 3 shows the station velocities of the DGFI solution in comparison with ITRF2000. There is in general a good agreement between both

Table 4: Comparison of space techniques at co-location sites. Shown are differences between local ties and computed station coordinates (upper part) and velocity differences between various techniques (lower part).

	GPS-VLBI	GPS-SLR	GPS-DORIS	SLR-VLBI
co-locations	35	24	21	12
missing ties	5	4	1	2
local tie differences				
< 5 mm	3	3	-	-
5 - 10 mm	6	5	-	2
10 - 20 mm	10	6	3	3
> 20 mm	11	6	17	5
velocity differences				
< 1 mm/yr	5	2	-	1
1 - 2.5 mm/yr	12	9	3	5
2.5 - 5 mm/yr	9	8	9	3
> 5 mm/yr	7	5	9	3

TRF realisations. Note that the combined TRF solution of DGFI contains less sites than ITRF2000, since we excluded a number of stations with short data time spans (e.g. < 1 yr), which in our opinion do not allow for accurate and reliable velocity estimations.

Table 5: Helmert transformation results between DGFI combined TRF solution and ITRF2000. For these transformations we used core reference stations for each technique.

H.-T. Results	VLBI	SLR	GPS	DORIS
Tx [mm]	1.1	-2.2	-0.4	1.4
Ty [mm]	-1.1	-0.3	-1.0	-0.2
Tz [mm]	0.5	-2.5	2.2	0.7
Scale [ppb]	0.15	0.23	-0.53	0.50
Tx rate [mm/yr]	0.3	0.0	-0.4	-0.5
Ty rate [mm/yr]	-0.1	-0.7	-0.1	-0.6
Tz rate [mm/yr]	-1.0	0.0	-1.0	-0.6
Scale rate [ppb/yr]	0.03	0.11	0.15	0.02
Pos RMS [mm]	2.2	4.0	1.6	12.6
Vel RMS [mm/yr]	0.51	0.85	0.90	2.12

4 Status of ITRF combination and deficiencies

The scope and accuracy of space geodetic observations, as well as the software systems, models and processing strategies have steadily been improved, and consequently remarkable progress could be achieved in the realisation of the TRF.

This paragraph summarises the status of the latest IERS realisation of the terrestrial reference frame, the ITRF2000 (Altamini et al., 2002). All ITRF2000 results and related files are available at http://lareg.ensg.ign.fr/ITRF/ITRF2000. Altamini et al. estimated the accuracy and long-term stability of the ITRF2000 scale and origin definition based on the contributing SLR and VLBI solutions. The WRMS values (propagated over 10 years) suggest a frame stability better than 4 mm in origin and better than 0.5 ppb in scale (equivalent to a shift of approximately 3 mm in station height). Furthermore the improvement of ITRF2000 compared to ITRF97 was shown. About 50% of station positions have an error less than 1 cm, and the velocities of about 100 sites have been determined at the 1 mm/yr level (or better). However, on the other hand, the ITRF2000 includes also a number of poorly estimated stations (e.g. mobile SLR and VLBI stations, GPS and DORIS stations with observation time spans less than one year). For about 20% of the ITRF sites the standard deviations for velocities are larger than 1 cm/yr, and for 5% the uncertainties exceed 10 cm/yr, with comparatively poor velocity estimations (see Figure 3).

The results of the external comparison of ITRF2000 with the combined DGFI solution demonstrate the high accuracy and consistency which have been achieved for the TRF realisations. On the other hand, it should be emphasised that despite this remarkable progress a number of shortcomings still exist. We have identified several deficiencies:

(1) Input data: The solutions submitted for ITRF2000 realisation include loose, minimum, or removable constraints. We found remarkable contradictions with respect to these declarations, and for some of the ITRF2000 input solutions the constraints could not be removed. If one neglects these constraints, significant biases and systematic effects might be introduced into the combination results.

(2) Systematic effects: Technique and/or solution-related systematic effects (biases) can be considered as one major limiting factor for the accuracy of the space geodetic observations and the combination results. The ITRF2000 results reveal that some of the contributing solutions differ considerably from the combined solution. The discrepancies reach up to 5 cm for the origin and a few ppb for the scale (Altamimi et al., 2002, see also: http://lareg.ensg.ign.fr/ITRF/ITRF2000/T.gif and D.gif). Recently published results (Angermann et al., 2002), and the comparison at co-location sites (see Table 3), show that the agreement between different techniques in several cases is not as good as it could be.

(3) ITRS products: Conventionally the ITRF is realised by the adoption of a set of positions and constant velocities. The observed non-linear effects in position time series (e.g. due to seismic or volcanic effects, deformations at plate boundary zones, local effects, etc.) conflict with the assumption of constant velocities (see Angermann et al., 2003). This can produce errors and systematic effects in the individual solutions, which may degrade the accuracy and consistency of the ITRF.

(4) IERS network: From our point of view, the current definition of the IERS network and the situation concerning co-location sites and local ties does not fully satisfy the accuracy, reliability and homogeneity requirements for a precise reference frame. As shown in Figure 2, the combined TRF solution includes only two sites with co-locations of all four techniques, and only ten sites with co-locations of VLBI, SLR and GPS. The results displayed in Table 4 show that the discrepancies between local ties and computed coordinates, and the velocity differences between techniques are unacceptably large in many cases.

(5) Combination methodology: A number of relevant aspects need to be studied in more detail, like for example, the level at which the combination should be performed (observation, normal equation or solution level), the weighting of solutions for the intra- and inter-technique combination, local ties and the equating velocities at co-location sites, the handling of non-linear site motions, and datum definition issues. In previous ITRF computations like the ITRF2000, different velocity estimations at co-location sites were forced to be identical. As can be seen in several cases the differences in velocities are quite large (see Table 4), and this may produce systematic errors in the combination results.

5 Recommendations for future ITRF combinations

Finally, we address some important aspects that should be considered for future ITRF realisations:

(1) Input data: To overcome the problems concerning the reduction of constraints, we recommend for future ITRF submissions that unconstrained normal equations. Then these normal equations can be combined directly without removing constraints. In the case of provided variance-covariance matrices it is important that the solutions are undeformed and that all constraints are completely reported.

(2) Systematic effects: To achieve further improvements there is an urgent need to investigate the existing differences between different techniques and/or solutions (e.g. modelling, parameterisation, software-related issues, analysis strategy, datum definition), and to understand their origin.

(3) ITRS products: Since various geophysical phenomena may vary as a non-linear function of time, the assumption of constant site velocities is not fulfilled. Consequently other products apart from station positions and (constant) velocities may be required, e.g. time series of position estimates. Major goals are to ensure the highest possible accuracy, reliability and long-term consistency of the frame.

(4) IERS network: The quality and reliability of stations and observations should be emphasised rather than the data quantity. It is essential to achieve improvements with regard to co-location sites and local ties, and to study their impact on the combination results.

(5) Combination methodology: Studies regarding various aspects (see previous section), and refinements of current combination methods are essential to achieve further improvements. The final goal is to develop optimal strategies for a rigorous combination of all techniques so as to ensure consistency between ITRS, ICRS and EOPs to the highest possible extent.

Acknowledgments. Thanks are expressed to Chris Rizos and an anonymous reviewer for their constructive comments. The research was supported by the programme Geotechnologien of BMBF and DFG, Grant 03F0336C; the publication no, is GEOTECH-39.

References

Altamimi, Z., P. Sillard, C. Boucher (2002). ITRF2000: A New Release of the International Terrestrial Reference Frame for Earth Science Applications. *J. Geophysical Res.*, 107 (B10), 2214, doi:10.1029/2001JB000561.

Angermann, D., K. Kaniuth, H. Müller, V. Tesmer (2002). Contribution of individual space techniques to the realization of vertical reference systems. In: *Vertical Reference Systems, IAG Symposia*, H. Drewes, A. Dodson, L. Fortes, L. Sanchez, P. Sandoval (eds), Vol. 124, pp. 91–96.

Angermann, D., B. Meisel, M. Krügel, H. Müller, V. Tesmer (2003). Time Evolution of the Terrestrial Reference Frame, presented at IUGG General Assembly, Sapporo.

Boucher, C., Z. Altamimi, P. Sillard (1999). The 1997 International Terrestrial Reference Frame (ITRF97). *IERS Technical Note 27*, Observatoire de Paris.

Ferland, R. (2002). Activities of the International GPS Service IGS Reference Frame Working Group. In: *Vistas for Geodesy in the New Millennium, IAG Symposia*, J. Adam and K.-P. Schwarz (eds), Vol. 125, pp. 3–8, Springer.

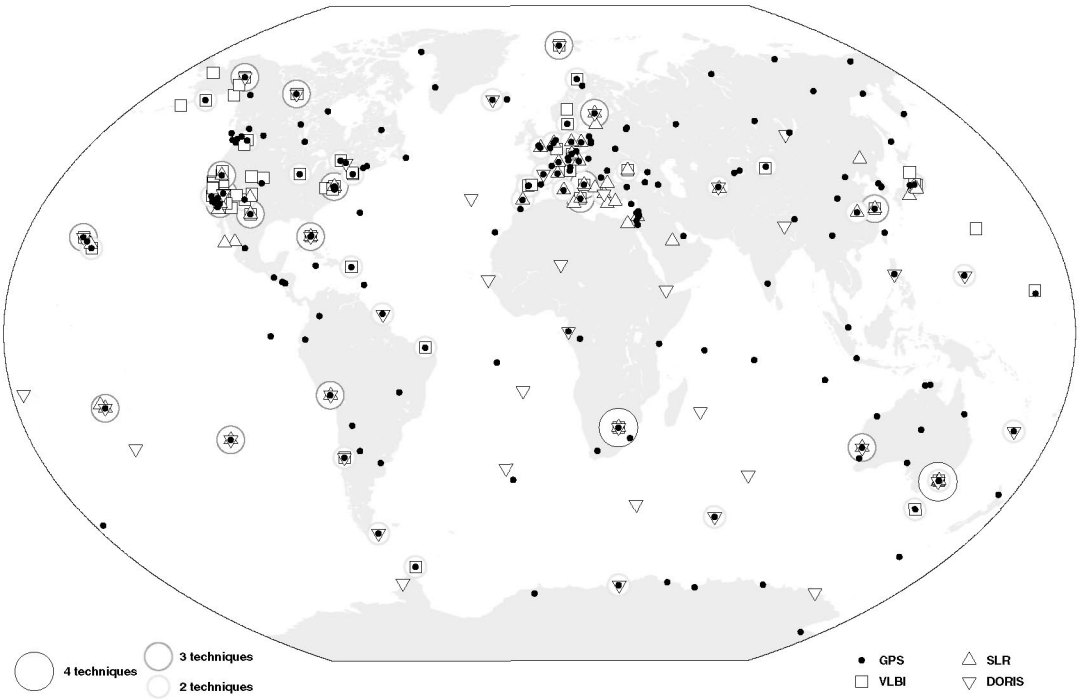


Fig. 2: VLBI, SLR, GPS and DORIS stations used for the TRF computation at DGFI. Furthermore co-location sites with two, three and four techniques are shown. Note that we excluded a number of poorly observed stations (e.g. less than one year of observations) from our combination.

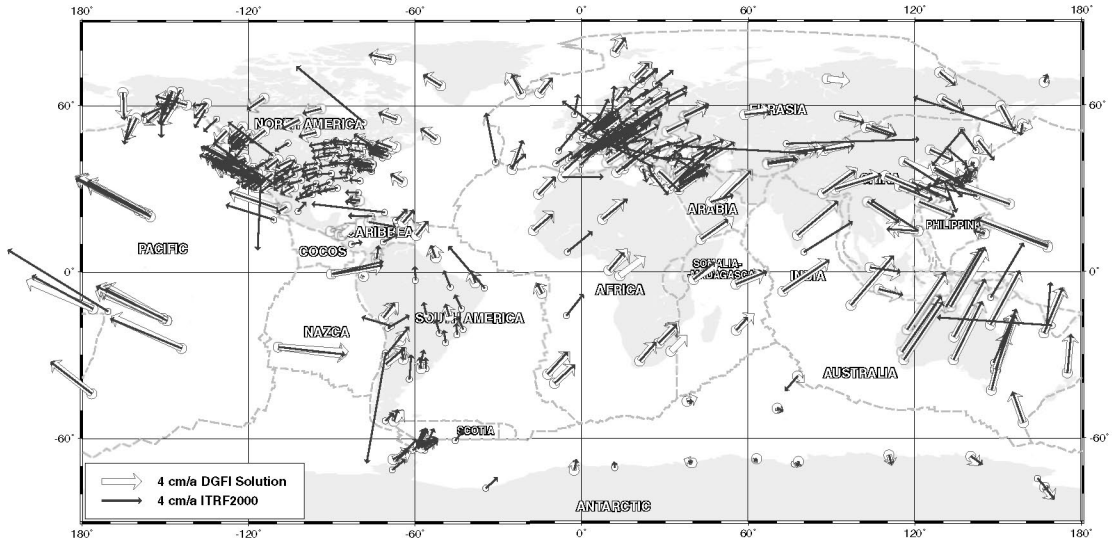


Fig. 3: Station velocites of DGFI solution compared to ITRF2000.

Time Evolution of the Terrestrial Reference Frame

D. Angermann, B. Meisel, M. Krügel, H. Müller, V. Tesmer
Deutsches Geodätisches Forschungsinstitut (DGFI), Marstallplatz 8, D-80539 München, Germany
email : angerman@dgfi.badw.de

Abstract. This paper investigates the time evolution of the terrestrial reference frame (TRF). We analysed time series of site positions and datum parameters obtained from VLBI, SLR, GPS and DORIS solutions with respect to non-linear motions (e.g. jumps, seasonal signals) and systematic differences. The time series of the translation and scale parameters were used to identify solution- and technique-related problems, and to investigate the contribution of the different techniques to realise the TRF origin and scale. The position time series reveal non-linear motions and jumps for many sites, which have to be considered for future TRF realisations.

Keywords. Terrestrial Reference Frame, ITRF, VLBI, SLR, GPS, DORIS, analysis of time series

1 Introduction and motivation

Future research, and in particular the new satellite missions (GRACE, JASON-1, ENVISAT and the future GOCE and GALILEO), will require a highly accurate and consistent global reference frame as a crucial basis to accomplish scientific goals, such as precise orbit determination, measurement of sea level change, postglacial rebound, plate tectonics, atmosphere dynamics, Earth orientation excitation or for other geodynamic and geophysical purposes. The quality of the realised reference frame has important implications for the ability to study these global and regional phenomena of the Earth's system, including the solid Earth, oceans and the atmosphere.

Since 1988, the International Earth Rotation and Reference Systems Service (IERS) has been in charge of the establishment and maintenance of an International Terrestrial Reference Frame (ITRF), a realisation of the International Terrestrial Reference System (ITRS). Conventionally, the ITRS is realised by the adoption of positions referred to a specific reference epoch and constant velocities for a set of global observing sites. The positions and velocities of these ITRF stations are derived from a combination of individual technique solutions provided by various analysis centers. The space geodetic techniques used at present are Very Long Baseline Interferometry (VLBI), Satellite and Lunar Laser Ranging (SLR/LLR), the Global Positioning System (GPS) and the Doppler Orbitography and Radio positioning Integrated by Satellite (DORIS). Due to the remarkable progress of these space geodetic observation techniques the accuracy and consistency of the ITRF has steadily increased. The latest realisation of the ITRS is the ITRF2000, which includes positions and velocities for about 800 stations located at approximately 500 sites (Altamimi et al., 2002).

In reality the assumption of constant site velocities is in conflict with non-linear effects caused by several geophysical phenomena (e.g. seismic or volcanic effects, quasi-periodic signals caused by loading effects). Not completely modelled site motions can also induce common (global) variations (e.g. translations, scale) of the station networks w.r.t. a crust fixed frame such as the ITRF2000. In addition, mass redistributions within the Earth system cause variations of the center of mass (geocenter). These time-varying effects on site positions and datum parameters can produce systematic errors in the individual solutions, which may also propagate into the ITRF and consequently degrade its accuracy.

In its function as IERS Combination Research Center and ITRS Combination Center, DGFI is deeply involved in the analysis and combination of different space geodetic observations. The results have shown that technique- and/or solution-related systematic effects (biases), which are often poorly characterised or quantified, are assumed to set the accuracy limits of the space geodetic observations in many cases. The analysis of time series is important for gaining insight into the solution characteristics and to identify technique- and/or solution-related problems. A comparison of position time series at co-location sites may help to separate and quantify possible effects, and to aid in understanding the origin of remaining discrepancies. This is of vital importance, so as to provide highly accurate and consistent results, and thus to exploit the full potential of the space geodetic techniques.

2 Data and processing strategy

The major characteristics of the contributing solutions are described below (summarised in Table 1).

VLBI solution: DGFI has re-analysed 2230 single 24-h VLBI sessions between January 1984 and the end of 2001. The computations were performed using the software OCCAM 5.0 (Titov et al., 2001). The models are based on the IERS Conventions 2000 (see http://www.iers.org/iers/products/conv/). For each daily VLBI session free normal equations with positions of the telescopes and the Earth Orientation Parameters were computed. The celestial reference frame was fixed to ICRF-Ext. 1 (e.g. Ma and Feissel, 1997). The datum was realised applying no-net-rotation (NNR) and -translation (NNT) conditions to all stations in each single session w.r.t. DGFI02R02 (the processing strategy is similar to that described in Tesmer, 2002).

SLR solution: The SLR results presented in this paper were obtained from a re-analysis of weekly combined LAGEOS-1 and 2 satellite arcs from November 1992 until February 2002 with the software package DOGS (DGFI Orbit and Geodetic parameter estimation Software). The processing strategy is similar to that described in the recently published results on the time evolution of an SLR reference frame (Angermann et al., 2002a). Due to the inhomogeneity of the SLR network, the number of stations and observations varies from arc to arc. In total 75 SLR stations observed the LAGEOS satellites during this period. We applied three condition equations to minimise the common rotation of the weekly SLR solutions w.r.t. ITRF2000. In addition we used weekly SLR solutions computed at ASI (Agencia Spaciale Italiano) using the software GEODYN. The ASI solutions are available at the data pool of the IERS SINEX Combination Campaign (see http://alpha.fesg.tu-muenchen.de/iers/sinex/data_pool.html).

GPS solutions: For the analysis of the GPS time series we used weekly solutions computed by three IGS Analysis Centers: CODE (Center for Orbit Determination in Europe), JPL (NASA Jet Propulsion Laboratory) and SIO (Scripps Institution of Oceanography). The weekly CODE and JPL solutions were accessed from the ftp server: ftp://igs.ensg.ign.fr/pub/igs/products, the weekly SIO solutions from: ftp://lox.ucsd.edu/pub/combinations. All these solutions are considered as loosely constrained w.r.t. ITRF2000 station positions. The time span for the GPS solutions covers six to seven years. The analysis centers included, in total, 150 to 180 stations, each week's solution containing between 50 to 150 stations. The GPS solutions were processed with different software packages: Bernese GPS Software (CODE), GIPSY (JPL) and GAMIT (SIO). For more information concerning the solution strategies of the institutions refer to the respective web pages: http://aiub.unibe.ch (CODE), http://igscb.jpl.nasa.gov (JPL) and ftp://lox.ucsd.edu/pub/combinations (SIO).

DORIS solution: For DORIS we used weekly SINEX files from IGN (Institut Géographique National) and JPL, available at the Crustal Dynamics Data Information System (CDDIS, ftp://cddisa.gsfc.nasa.gov/pub/doris/products/sinex_series/ignwd03). The solutions were computed using the GIPSY/OASIS software, and data from October 1992 to May 2002 for all DORIS satellites. We used those solutions (version 3) that were obtained applying a free network approach, and afterwards they were rotated to ITRF2000 (see description file: ign.wd03.snx.dsc)

Datum definition: In order to realise the datum of the individual VLBI, SLR, GPS and DORIS solutions in a similar way, we applied 7 parameter Helmert transformation models (3 translations, 3 rotations, 1 scale) to transform them into a consistent reference frame, the ITRF2000. A well-known problem related to similarity transformations is the fact that the results are sensitive to the geometric distribution and to changes w.r.t. the selected transformation stations. To minimise these effects, we used a subset of "good" and well distributed stations for each technique (so-called reference stations).

In the case of VLBI typically only 4-6 telescopes observe simultaneously within one daily session, and the station configuration often changes from one session to the other. The SLR data are relatively inhomogeneously distributed in space and time. We defined a subset of 20 reference stations with dense tracking and a good global distribution. For the transformation of the weekly GPS solutions we used the 55 reference stations identified by the IGS (Ferland, 2002). For the transformation of the DORIS solutions we used 50 reference stations evenly distributed across the globe. Similar to GPS, the effect of possible changes of the station configuration can be expected to be small.

3 Time evolution of the individual reference frames

To gain an insight into the characteristics and the contribution of the individual space techniques to the realisation of the terrestrial reference frame, we analysed the time series of scale and translation variations derived from the 7 parameter Helmert transformation models between the individual space technique solutions and ITRF2000. The time series of these parameters reflect common (global) effects of the station networks w.r.t. ITRF2000.

Figure 1 shows the time series of scale variations. The VLBI scale variations of the daily session solutions have a higher noise level compared to the weekly solutions of the other techniques, mainly due to the poor network geometry of single VLBI sessions and due to the fact that the solutions span only one day. In principle, the VLBI and SLR scale is in good agreement with ITRF2000 scale, except for some irregularities in the ASI solution in early 1999. The DORIS scale has an offset of about 4 ppb w.r.t. ITRF2000. The three GPS series (CODE, JPL and SIO) agree well (within 1 ppb) during the last two years, whereas before 2000 larger discrepancies and some irregularities exist. The significant jump of about 2 ppb in the SIO scale in early 2000 was probably caused by a change of the elevation cut-off angle (Herring, 2002). The large scale variations of the daily VLBI session solutions can be smoothed significantly by a monthly accumulation (see Figure 2).

Through the orbit dynamics satellite methods such as SLR, GPS and DORIS are sensitive to the center of mass (geocenter). Therefore these satellite techniques can be used to realise the origin of the terrestrial reference frame. The time series of weekly translation (origin) variations w.r.t. ITRF2000 derived from the individual solutions are shown in Figure 3. The most stable results were obtained from SLR. The amplitudes for the annual translation variations of both SLR solutions are almost identical (see Figure 4). There is also a good agreement with recently published SLR results (Angermann et al., 2002b) and with GPS-derived results (e.g. Dong et al., 2003). The time series for the translation parameters derived from the GPS and DORIS solutions show larger variations compared to SLR, in particular for the z-component. DORIS shows significant annual signals in the x-and y-component, but GPS only in the y-component. The large z-offset of DORIS in 1998 was caused by SPOT-4 data problems (Willis, 2003).

4 Site position time series

We analysed the position time series obtained from the different space technique solutions with respect to non-linear effects, periodic signals and systematic differences (biases), and compared the results at co-location sites. For a large number of tracking sites we observed non-linear effects in the position time series that conflict with the assumption of constant velocities.

For many of the tracking stations, especially for those with long data time series, the instrumentation changed several times due to system upgrades, antenna and receiver changes, etc. The effects of instrumental changes are illustrated in Figure 5. At GPS station Onsala a change of the radome in early 1999 caused a jump of about 2 cm in the height component (Kaniuth and Huber, 2003). Several receiver and antenna changes at GPS station Westford are accompanied by significant jumps in the longitude. Both co-location sites do not show similar effects in the VLBI position series, indicating that the observed jumps are a technique-related problem, and not a "real" site motion caused by geophysical phenomena.

During the last few years the software systems, models and processing strategies have improved significantly. To achieve consistent results it is necessary to reprocess all data with the latest software version, state-of-the-art models and the same strategy. In the case of VLBI and SLR all data were reprocessed in a consistent way. At present, this is in particular a problem in the case of GPS, as it requires a tremendous effort to reprocess all GPS data homogeneously. As a consequence many of the time series are affected by changes due to inconsistent software, models and processing strategy (e.g. Rothacher, 2002). An example is the jump in the time series of the SIO scale, caused by a change of the elevation cut-off angle (see Figure 1).

A number of stations are located in deformation zones, which cover about 15% of the Earth's surface. In these geodynamically active regions large earthquakes occur quite often, which are accompanied by seismic deformation processes that may cause non-linear motions of surrounding stations. In the last few years numerous GPS projects were initiated to study the deformations in plate boundary zones. The effect of large earthquakes on the position time series is illustrated for three stations (see Figure 6). Earthquakes in Arequipa, Peru in June 2001 caused a jump of about 50 cm horizontally (Kaniuth et al., 2002), and the motion of this station after the earthquake differs significantly from the expected long-term mo-

tion. This change in motion is probably caused by post-seismic relaxation processes, which also have been observed after the 1995 Mw 8.0 Antofagasta earthquake in Northern Chile (Klotz et al., 1999, Klotz et al., 2001). For two other stations displayed in Figure 6 (Ankara and Cocos Island) earthquakes caused significant jumps of a few cm in the position time series, accompanied by site motions after the earthquakes that are different from their long term behaviour.

Site displacements are also caused by solid Earth tides, ocean tidal loading and atmospheric loading, etc. In principle, these effects can mostly be corrected by models (see IERS Conventions 2000, http://www.iers.org/iers/products/conv/). In practice, there are still deficiencies in the models, which can cause quasi-periodic variations in the position time series. Precise positioning of globally distributed GPS sites reveals that during February to March, the Northern Hemisphere compresses (and the Southern Hemisphere expands), so that sites near the North Pole move downward by 3.0 mm, and sites near the equator move northwards by 1.5 mm. As a response of an elastic Earth to increased winter loading of soil moisture, snow cover, and atmosphere, the opposite pattern of deformation occurs during August and September (Blewitt et al., 2001).

5 Conclusions

The time series of the datum parameters and site positions are influenced by remaining technique- and solution-related problems, that have to be separated from "real" geophysical signals. The results confirm that SLR is the best technique to realise the origin of the terrestrial reference frame. VLBI and SLR realise the scale consistent with ITRF2000. Improvements to the GPS and DORIS solutions seem to be necessary in order to use them for TRF datum realisation.

The observed non-linear effects in the position time series are in conflict with the assumption of constant site velocities. This can produce errors and systematic effects (biases) in the individual solutions, which may degrade the accuracy and consistency of the ITRF. Hence a better realisation of the terrestrial reference frame may require the estimation of non-linear components in site positions.

Acknowledgments We thank the reviewers Chris Rizos and Pascal Willis for their constructive comments. The research was supported by the programme Geotechnologien of BMBF and DFG, Grant 03F0336C; the publication no. is GEOTECH-38.

References

Altamimi, Z., P. Sillard, C. Boucher (2002). ITRF2000: A New Release of the International Terrestrial Reference Frame for Earth Science Applications. *J. Geophysical Res.*, 107 (B10), 2214, doi:10.1029/2001JB000561.

Angermann, D., M. Gerstl, R. Kelm, H. Müller, W. Seemüller, M. Vei (2002a). Time Evolution of SLR Reference Frame. *Adv. in Space Res.*, Vol. 30/2, pp. 201–206, Elsevier.

Angermann, D., H. Müller, M. Gerstl (2002b). Geocenter variations derived from SLR data to LAGEOS 1 and 2. In: *Vistas for Geodesy in the New Millennium, IAG Symposia*, J. Adam and K.-P. Schwarz (eds), Vol. 125, pp. 30–35.

Blewitt, G. D., Lavallee, P. Clarke, K. Nurutdinov (2001). A new global mode of Earth deformation: Seasonal cycle detected. *Science*, 294 (5550), pp. 2342–2345.

Blewitt (2003). Self-consistency in reference frames, geocenter definition, and surface loading of the solid Earth. *J. Geophys. Res.*, Vol. 108, No. B2, 2103, doi: 10.1029/2002JB002082.

Dong, D., T. Yunck, M. Heflin (2003). Origin of the International Terrestrial Reference Frame. *J. Geophys. Res.*, Vol. 108, No. B4, 2200, doi: 10.1029/2002JB002035.

Ferland, R. (2002). Activities of the International GPS Service IGS Reference Frame Working Group. In: *Vistas for Geodesy in the New Millennium, IAG Symposia*, J. Adam and K.-P. Schwarz (eds), Vol. 125, pp. 3–8, Springer.

Herring, T. (2002). Personal Communication.

Kaniuth, K., and S. Huber (2003). An assessment of radome effects on height estimates in the EUREF network. *Mitt. Bundesamt für Kartographie und Geodäsie*, 29, pp. 97-102.

Kaniuth, K., H. Müller, W. Seemüller (2002). Displacement of the space geodetic observatory Arequipa due to recent earthquakes. *Zeitschrift für Vermessungswesen*, Heft 4, pp. 238–243, Witwer.

Klotz, J., D. Angermann, G.W. Michel, R. Porth, C. Reigber, J. Reinking, J. Viramonte, R. Perdomo, V.H. Rios, S. Barientos, R. Barriga, O. Cifuentes (1999). GPS-derived deformation of the Central Andes including the 1995 Antofagasta Mw=8.0 Earthquake. *Pure and Appl. Geophysics*, 154, pp. 709–730.

Klotz, J., G. Khazaradze, D. Angermann, C. Reigber, R. Perdomo, O. Cifuentes (2001). Earthquake cycle dominates contemporary crustal deformation in Central and Southern Andes. *Earth and Planetary Science Letters*, 193, pp. 437–446, Elsevier.

Ma, C., M. Feissel (1997). Definition and Realisation of the International Celestial Reference System by Astrometry of Extragalactic Objects. *IERS Technical Note 23*, Paris.

Rothacher, M. (2002). Estimation of station heights with GPS. In: *Vertical Reference Systems, IAG Symposia*, H. Drewes, A. Dodson, L. Fortes, L. Sanchez, P. Sandoval (eds), Vol. 124, pp. 81–90, Springer.

Tesmer, V. (2002). VLBI Solution DGFI01R01 based on least-squares estimation using OCCAM5.0 and DOGS-CS. In: *IVS 2002 General Meeting Proceedings*, N. Vandenberg and N.K. Baver (eds), NASA/CP-2002-210002.

Titov, O., V. Tesmer, J. Böhm (2001). OCCAM5.0 Users Guide, *AUSLIG Technical Report 7*, Canberra.

Willis, P. (2003). Personal Communication.

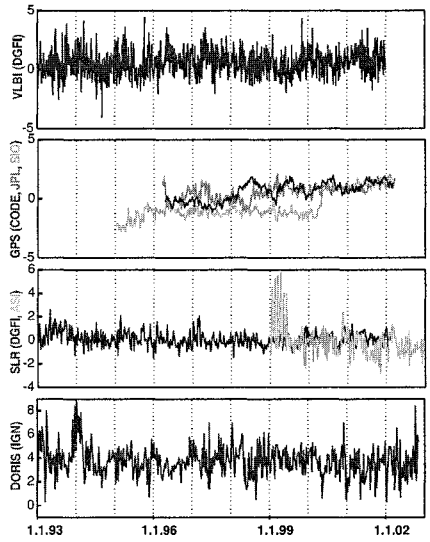


Fig. 1: Time series of scale variations [ppb].

Table 1: Summary of solutions used for this study.

Technique	Analysis Center	Software	Data Time Span	Number of Stations
DORIS	IGN/JPL	GIPSY / OASIS	1992.8-2002.8	80
GPS	CODE	Bernese	1996.0-2002.2	163
GPS	JPL	GIPSY	1996.0-2002.2	180
GPS	SIO	GAMIT	1995.0-2002.2	157
SLR	ASI	GEODYN	1999.0-2003.0	35
SLR	DGFI	DOGS	1992.9-2002.2	75
VLBI	DGFI	OCCAM	1984.0-2002.0	56

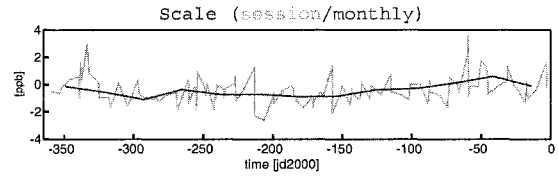


Fig. 2: VLBI scale variations obtained from daily session and accumulated monthly solutions. The normal equations of the daily sessions for the year 1999 were accumulated monthly with the DGFI combination software DOGS-CS.

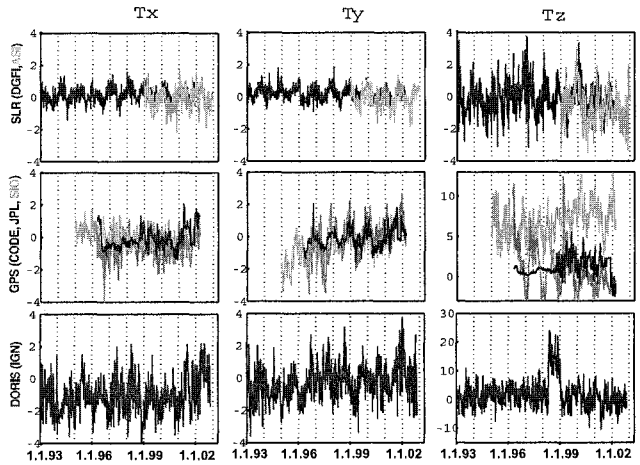


Fig. 3: Time series of weekly translation variations [cm].

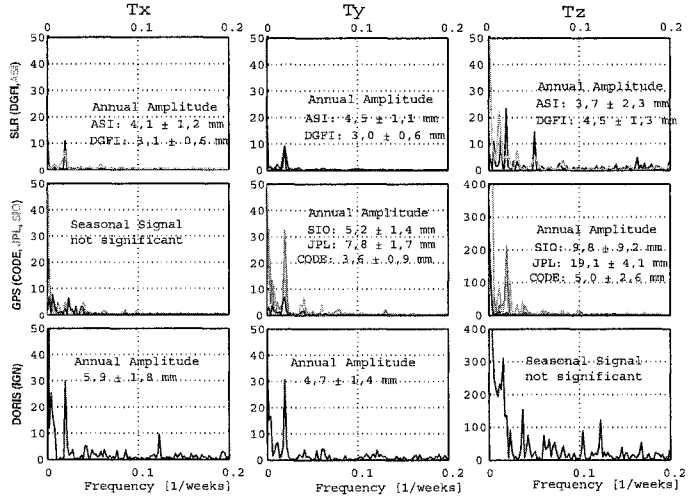


Fig. 4: Fourierspectrum of the translation parameter time series

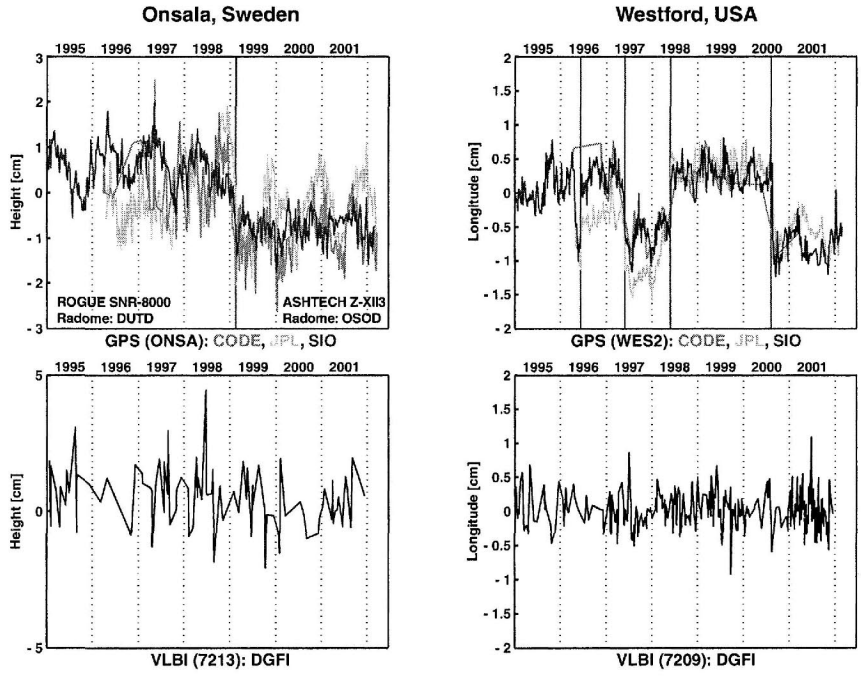


Fig. 5: The effect of instrumental changes on the position time series is shown for two co-location sites.

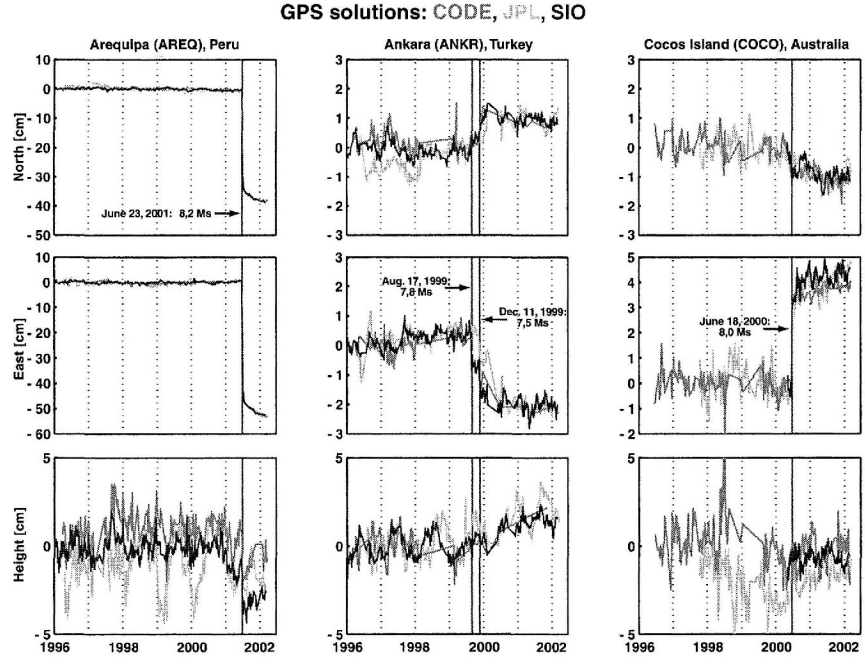


Fig. 6: The effect of large earthquakes on the position time series is shown for three stations.

Scale variation of GPS time series

Markku Poutanen, Jorma Jokela, Matti Ollikainen, Hannu Koivula, Mirjam Bilker, Heikki Virtanen
Finnish Geodetic Institute
Geodeetinrinne 2, FIN-02430 Masala Finland
Markku.Poutanen@fgi.fi

Abstract. We give an overview of time series analyses of permanent GPS stations using solutions of the IGS and FinnRef® networks. Lomb periodograms show in most cases a statistically significant annual period both in station coordinates and inter-station distances. In regional networks the scale of the whole network changes periodically, and in some cases there is also a secular trend. There are several possible causes of scale variations, which may not be separable in the data. These include computational artefacts, periodic systematic errors in satellite orbits, signal path delay variations, and geophysical causes like loading and postglacial rebound. We discuss possible reasons, their significance, and their consequences on high-precision GPS observations. Additional constraints, e.g. time series from the superconducting gravimeter, are also discussed.

Keywords. GPS time series, periodic variations

1 Introduction

The Finnish permanent GPS network FinnRef® consists of 13 GPS stations. The network is the backbone of the Finnish realisation of the EUREF frame, referred to as EUREF-FIN. Four stations in the FinnRef network belong to the EUREF permanent GPS-network (EPN), and one station belongs to the network of the International GPS Service (IGS). Through these stations FinnRef® creates a seamless connection to the global reference frames. The observations obtained by FinnRef® stations can be used for studying the geodynamics of the Earth's crust (Fig. 1).

All FinnRef® stations are also used in the computation of the joint Nordic GPS network in the BIFROST (Baseline Inferences for Fennoscandian Rebound Observations, Sea Level and Tectonics) project (Milne *et al.*, 2001, Scherneck *et al.*, 2002).

Currently we have a seven year GPS time series. Quite soon after the beginning it became obvious that there were both seasonal and short term variations. Because one of the tasks was to measure the land uplift rate, we had to take the variation into account. In the worst case it could lead to poor results or even misinterpretation, as demonstrated in Fig. 2.

The variation is visible both in the vertical component and in inter-station distances. In distances the amplitude of the variation depends on the vector length, thus leading us to consider the phenomenon as a possible scale error.

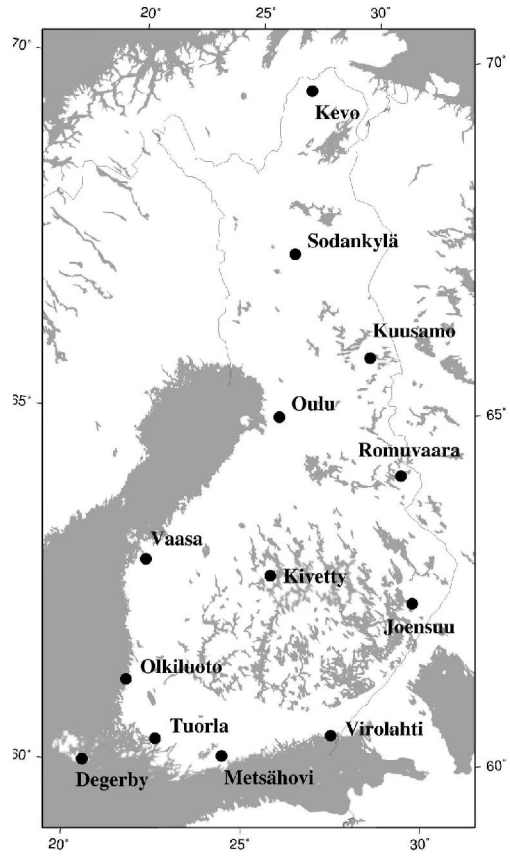


Fig. 1. Finnish permanent GPS network, FinnRef®. Subsets of points belong to the IGS and EPN networks, and all points are used for Nordic geodynamics studies. FinnRef® is additionally used as a fundamental network for the definition of the new Finnish reference frame EUREF-FIN (Ollikainen *et al.*, 2000).

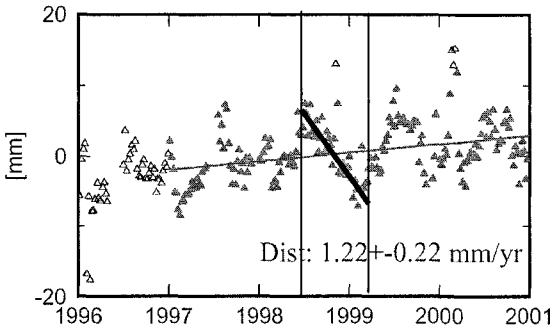


Fig. 2. An example of a possibility for a misinterpretation. The actual secular change in the distance is 1.2 mm/yr. If one takes a 9-month sequence, the annual variation may result in an unrealistic –10 mm/yr.

2 Environmental effects

Our first discovery, concerning the seasonal effect is shown in Fig. 3, where the height of the Sodankylä station changes 20-40 mm during the winter time. This was soon attibuted to the accumulation of snow on top of the antenna radome. We see the effect on two FinnRef® stations; on the others, the snow causes only a minor temporary variation, if any. A similar effect can also be seen on some northernmost Swedish permanent stations. This is discussed e.g. in Johansson *et al.* (2002) and Mäkinen *et al.* (2003).

In solving for trends from the time series, we have used different methods to deal with data points corrupted by the snow effect. We fitted iteratively for the trend with least squares, at each round giving zero weight to the points with residuals larger than a pre-selected cutoff value. This value is not very critical for the final result; we used twice the root-mean-square scatter in the summer solutions. In the second method we simply rejected all winter data in our solution thus rejecting also some good data. One more trial was done fitting simultaneously a trend and a sinusoid to the observations. A comparison between methods and a more detailed analysis of the rates are discussed in Mäkinen *et al.* (2003) and Poutanen *et al* (2002).

It is evident that removing some data will lead to data gaps and it may cause bias in our trend determination. Because the number of rejected points can vary from year to year and station to station, the seasonal data is unequally present at different stations and in different years. Also, some values may be only slightly corrupted, thus escaping our rejection criteria but still causing a bias in solutions.

In some the earliest BIFROST solutions the uncorrected snow bias caused unrealistic high uplift values for the Kuusamo station, suggesting a shift of the whole land uplift maximum in a wrong place. It turns out that removing environmental effects is not trivial, and requires special attention.

3 Annual and short term scale variations

The time series resulting from our weekly GPS solutions show periodic variation, which was noticed for almost all baseline components of the FinnRef® network. Because the baseline length is less noisy and the most sensitive quantity to show the changes, and is free from changes in the ITRF reference frame, we concentrated on investigating these lengths. Additionally, using the double difference solution of the Bernese software, this was the most natural observable.

The existence of the annual period was confirmed by the Lomb periodigram (Press et al., 1996) shown in Fig. 4. The annual periodic variation was found in practically all baselines of the FinnRef® network. The periodic variation in the baseline length is in the same phase and the amplitude depends linearly on the baseline length (Fig. 5). This happens only if the scale of the whole network is changing.

In order to find out if the scale of the network has a periodic variation we solved for the 7-parameter Helmert transformations between the first and the other weekly solutions of the FinnRef® coordinates. The scale factors of the weekly solutions indicated by the Helmert transformations show an annual variation (Fig. 6).

The amplitude of the annual period is 4 ppb, i.e. it corresponds 4 mm change for a 1000 km long baseline. This can be explained by the change of height of the whole network. One-centimetre change in network height causes a 1.5 ppb scale error in the network. The observed scale variation can be explained by a 3-4 cm height change of the network.

Such a height change can be explained by the loading variation. The seasonal variation can be caused by the atmospheric pressure changes, snow cover, soil moisture or ocean water masses. The unmodelled loading effects can cause periodic, but also short-term, changes in the GPS time series. Many of these are visible in all components, but especially in height and scale.

In addition to the seasonal signal, short-term variations are also visible (Poutanen *et al.*, 2002). The observed variations can be attributed to the weather cycle and loading due to the varying air pressure. A correlation with the observed air pressure data is reported in Poutanen *et al.* (2002).

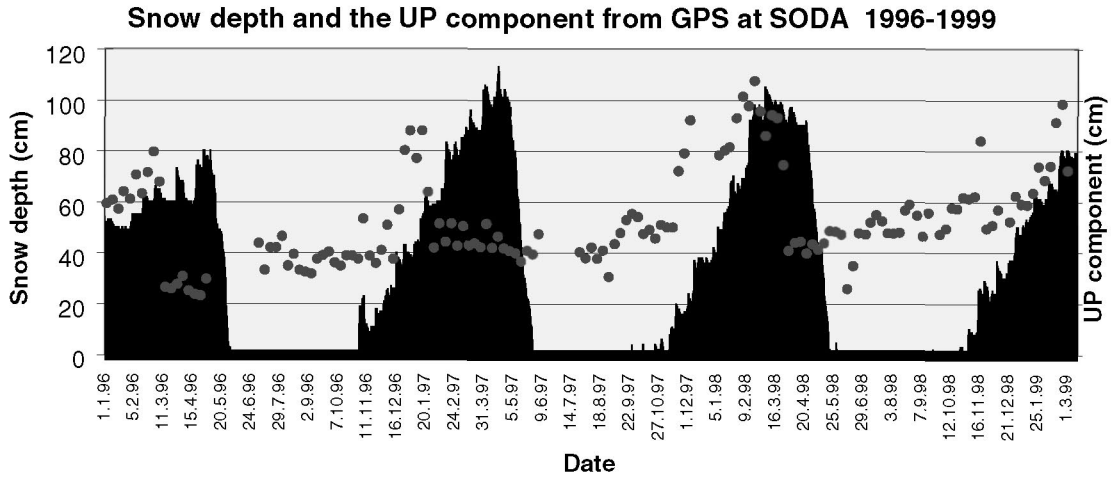


Fig. 3. An example of the environmental effect. Dots represent the height component of the Sodankylä GPS station. In winter time the height component may change 2-4 cm. The black area shows the depth of the snow. The antenna height error does not fully correlate with snow depth. The reason is that the snow accumulated on top of the antenna falls down when the temperature rises above zero.

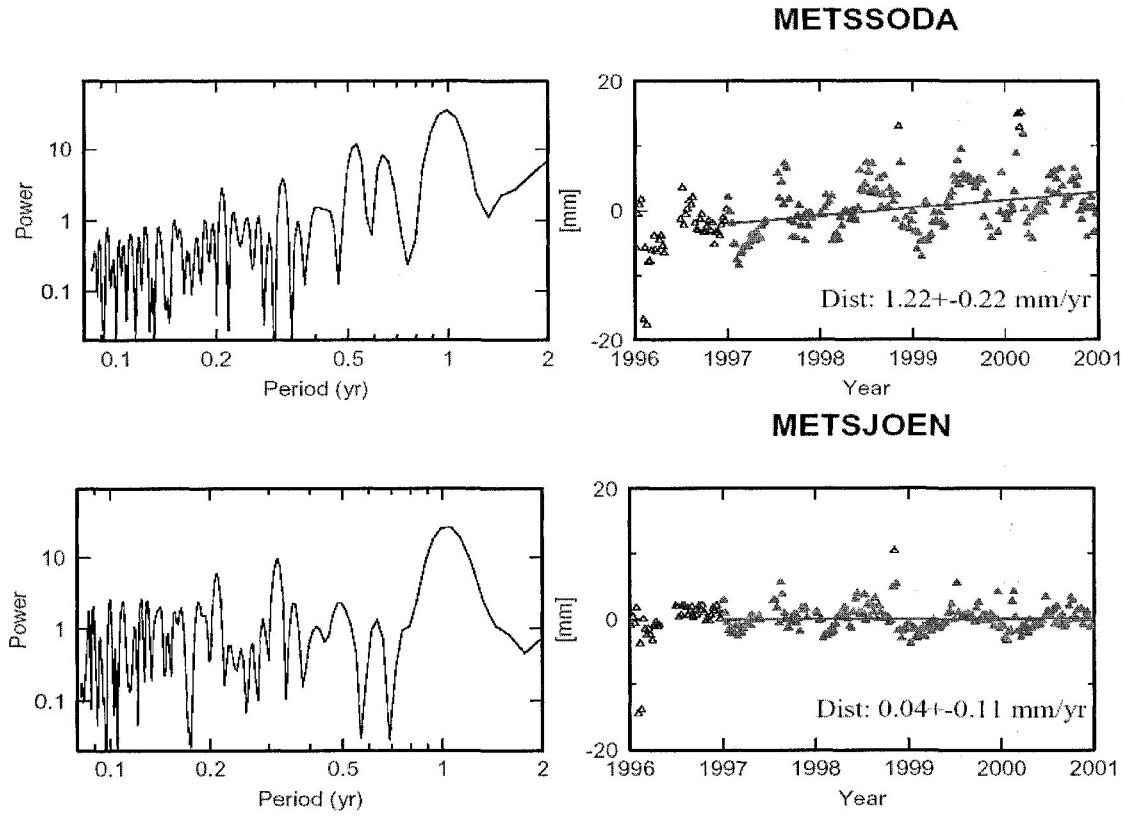


Fig. 4. Two examples of GPS time series and their Lomb periodograms, expressed in the time domain. In the upper plot the distance between stations (Metsähovi – Sodankylä) is about twice the distance in the lower plot (Metsähovi – Joensuu). In both cases the annual peak in the Lomb periodogram (left) is statistically significant.

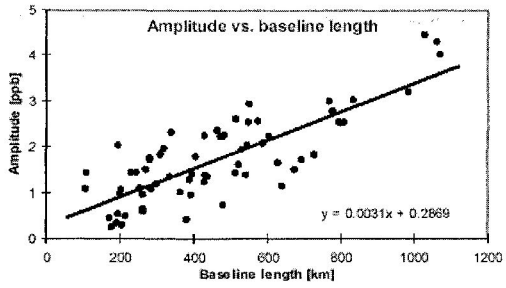


Fig. 5. Scale change in the Finnish permanent GPS network FinnRef®. The amplitude of the annual variation depends on the distance, so the scale change is about 3-4 ppb.

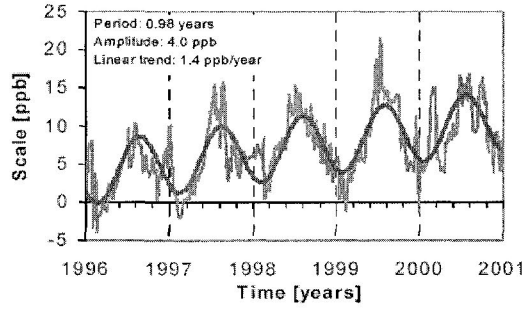


Fig. 6. Annual variation and a secular trend in the FinnRef® network.

4 Global and regional scale change

The FinnRef® annual scale variation is in the same phase as the variation of the DORIS positioning system results at the Metsähovi observatory (Mangiarotti et al. 2000). In the DORIS time series we have an indication of phase shift between Northern and Southern hemisphere. Global loading of atmosphere, snow and other seasonal varying phenomena has been suggested as a reason for this, see van Dam *et al.* (2001).

The IGS GPS time series shows the same effect, both in the radial component and in inter-station distances. In Fig. 7 we give two examples of IGS time series, one in the Northern hemisphere (IRKT in Siberia), and the other one in the Southern hemisphere (YAR1 in Australia) where the phases are opposite.

More stations are shown in Fig. 8. Generally, the variations in the Northern hemisphere have their maximum in summer or autumn, whereas for the Southern hemisphere stations the maximum phases occur in the first quarter of the year. The same behaviour can also be seen in the interstation distances.

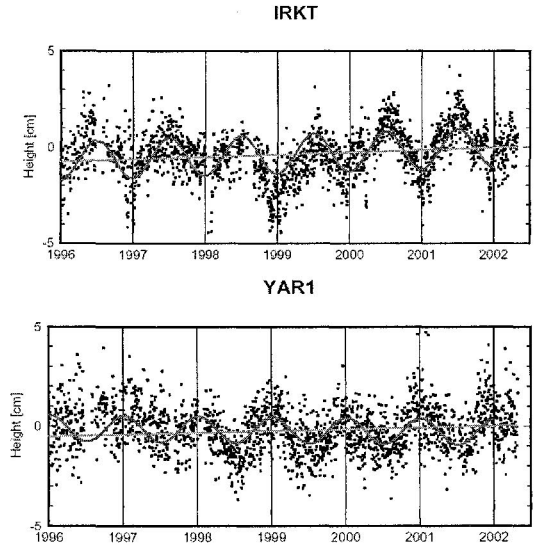


Fig. 7. Time series of two IGS stations taken from the IGS ppp solution. IRKT is in the Northern hemisphere, YAR1 in the Southern hemisphere. The phase of the annual variation is opposite.

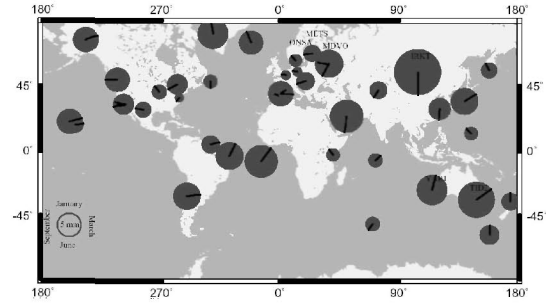


Fig. 8. Amplitude and phase of radial variation of some IGS GPS stations. The direction of the arrow shows the time of the maximum phase and the size of the circle is the amplitude.

We also studied the scale behaviour of the IGS GPS network solution. Because the annual variation is in different phases globally, we see no periodicity. The scale in the IGS network is quite constant at a global level (Fig. 9, topmost panel). However, in the Fennoscandian area (Fig. 9, middle panel), there is a secular change, as is seen in the FinnRef® data also. This supports our hypothesis that the postglacial rebound is the reason for this secular change.

Fig. 9 shows also one additional problem connected to the use of the ITRF. The scale of the network is not constant, but it has regional deficienceis because of unmodelled vertical crustal motions.

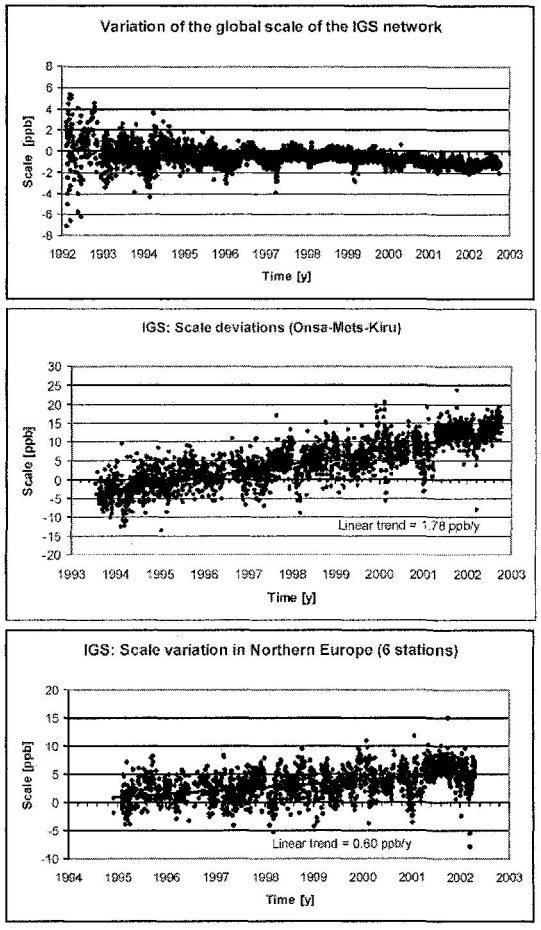


Fig. 9. Scale change in the IGS network. Globally the scale change is quite small and there is only a minor trend (top). Three Nordic stations show a trend of about 1.8 ppb/yr (middle). Six North European stations show much smaller trend (bottom). This indicates that the effect of the postglacial rebound in the GPS network scale is visible.

5 Additional observations

In Fig. 10 we show an example of detecting atmospheric loading using GPS. There are two phenomena which are not separable in GPS data if no additional information is available. First, the changing air mass and humidity causes a change in the tropospheric path delay, and secondly, the crust is displaced by varying loading. Both phenomena can be in cm-range and thus well within the detection limit of GPS observations.

In connection to the data reduction for the Metsähovi superconducting gravimeter, we have calculated the atmospheric loading from HIRLAM (High Resolution Limited Area Model) air pressure data and sea loading from tide gauge data. Loading calculations of the Baltic Sea show that 1 m of uniform layer of water corresponds to 31 nm s^{-2} in gravity and 11 mm in height.

Validity of both calculations has been confirmed by data from the superconducting gravimeter GWR T020 at Metsähovi. The range of air pressure variation at Metsähovi is 100 hPa. The vertical motion due to air pressure is about four times larger than the loading effect due to the Baltic Sea. Simultaneous regression with the air pressure and the Baltic Sea level in Helsinki gives vertical motion –0.35 mm/hPa and –12.4 mm/m (Virtanen, 2002).

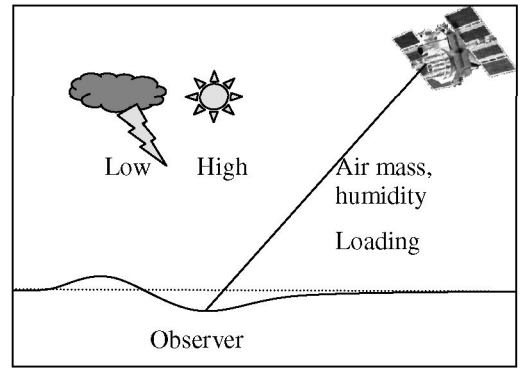


Fig. 10. Atmospheric phenomena on GPS observations. Air pressure variations cause a loading effect that can be up to 3-4 cm. Additionally, air mass and humidity changes affect the tropospheric path delay. These are not separable in the GPS observations.

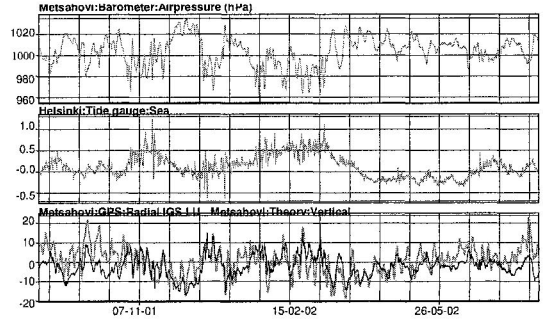


Fig. 11. (Top) Air pressure at Metsähovi [hPa] from September 2001 to July 2002. (Middle) Sea level [m] at the Helsinki tide gauge. (Bottom) IGS radial component [mm] of Metsähovi GPS station (black) and calculated vertical motion due to the loading (gray).

In Fig. 11 we show the air pressure variation at Metsähovi during one year, and the sea level variation at the Helsinki tide gauge, about 30 km

from the station. The loading computed from this data and the vertical component of the IGS GPS solution show good correlation in general trends, but many details need still further study. Direct comparison with superconducting gravimeter data is yet to be done.

6 Conclusions

We have discussed various scale-related phenomena in GPS time series. It is obvious that the temporal variation degrades our resolution and should be taken into account in high-precision GPS observations.

GPS observations alone do not solve the problem. We need data from other sources, like air pressure and tide gauge observations, to compute the actual loading on site. A superconducting gravimeter will give an independent constraint to the vertical motion. This concept is compatible with the IGGOS and ECGN proposals to establish multi-instrument reference networks (Ihde, 2003, Rummel *et al.*, 2001). In the Nordic countries a similar proposal for a NGGOS (Nordic Geodetic and Geodynamic Observing System) has been developed (Poutanen *et al.*, 2003).

New applications, like precise point positioning technique or meteorological applications to estimate the tropospheric water vapour content require even sub-cm accuracy in the satellite range measurements. Our examples show that this is not possible if actual loading is not taken into account. Also regional scale variation requires more careful consideration of secular crustal movements when determing the ITRF reference frame in the future.

References

Johansson J.M., J.L. Davis, H.-G. Scherneck, G.A. Milne, M. Vermeer, J.X. Mitrovica, R.A. Bennett, M. Ekman, G. Elgered, P. Elosegui, H. Koivula, M. Poutanen, B.O. Rönnäng, and I.I. Shapiro (2002): Continuous GPS measurements of postglacial adjustment in Fennoscandia, 1. Geodetic results. *J. Geophys. Res.* **107**, B8.

Ihde J. 2003. European Combined Geodetic Network (ECGN). 1st call for participation. Implementation of the ECGN stations.

Mäkinen J., H. Koivula, M. Poutanen, V. Saaranen (2003): Vertical velocities in Finland from permanent GPS networks and from repeated precise levelling. *Journal of Geodynamics*, **38**, 443-456.

Mangiarotti, S., A. Cazenave, L. Soudarin, J.-F. Cretaux (2000): Annual vertical crustal motions predicted from surface mass redistribution and observed by space geodesy, *J. Geophys. Res.*, Vol. **106**, No. B3, p. 4277-4291.

Milne G.A., J.L. Davis, J.X. Mitrovica, H.-G. Scherneck, J.M. Johansson, M. Vermeer and H. Koivula (2001): Space-Geodetic Constraints on Glacial Isostatic Adjustment in Fennoscandia. *Science*, Vol. **291**, p. 2381-2385.

Ollikainen, M., H. Koivula and M. Poutanen (2000): The Densification of the EUREF Network in Finland. IAG, Section I - Positioning, Commission X - Global and Regional Geodetic Networks, Sub-Commission for Europe (EUREF). Report on the Symposium of the IAG Subcommission for Europe (EUREF) held in Prague, 2-5 June 1999. *Veröffentlichungen der Bayerischen Kommission für die Internationale Erdmessung der Bayerische Akademie der Wissenschaften, Astronomisch-Geodätische Arbeiten*, Heft Nr. **60**. München. p. 114-122.

Poutanen, M., H. Koivula, M. Ollikainen (2002): On the periodicity of GPS time series. International Association of Geodesy Symposia, Vol. **125**. *Vistas for Geodesy in the New Millennium.* (Eds J. Ádám and K.-P. Schwarz). IAG 2001 Scientific Assembly, Budapest, Hungary, September 2-7, 2001. pp. 388-392. Springer-Verlag Berlin Heidelberg New York.

Poutanen M., H.-P. Plag, H.-G. Scherneck, P. Knudsen, M. Lilje (2003): NGGOS - Nordic Geodetic and Geodynamic Observing System. *A paper presented to the Presisium of the Nordic Geodetic Commision,* May 30, 2003.

Press W.H., S.A. Teukolsky, W.T.Vetterling, and B.P. Flannery (1996): Numerical Recipes in Fortran 77: The art of scientific computing, Second Edition, Cambridge University Press, ISBN 052143064X.

Rummel, R., Drewes, H., and Beutler, G., 2001. Integrated Global Geodetic Observing System (IGGOS): A candidate IAG project, presented at IAG General Assembly, Budapest, 2001.

Scherneck, H.-G., J.M. Johansson, G. Elgered, J.L. Davis, B. Jonsson, G. Hedling, H. Koivula, M. Ollikainen, M. Poutanen, M. Vermeer, J.X. Mitrovica, and G.A. Milne (2002): BIFROST: Observing the Three-Dimensional Deformation of Fennoscandia. In *Ice Sheets, Sea Level and the Dynamic Earth*, (Eds. J.X. Mitrovica and B.L.A. Vermeersen). American Geophysical Union, Geodynamics Series, Volume **29**, Washington, D.C., p. 69-93.

van Dam T., J. Wahr, P.C.D. Milly, A.B. Shmakin, G. Blewitt, D. Lavallee, and K.M. Larsson (2001): Crustal displacement due to continental water loading. *Geophysical Research Letters*, Vol. **28**, No, 4, p. 651-654.

Virtanen, H. (2002): Air pressure and Baltic Sea loading corrections for gravity data at Metsähovi. *Proceedings of the XIV General Meeting of the Nordic Geodetic Commission* (Ed. M. Poutanen, H. Suurmäki). Finnish Geodetic Institute, Kirkkonummi.

Vertical repeatability of GPS measurements in the Tsukuba GPS/MET dense-network campaigns

S. Shimada

Solid Earth Research Group, National Research Institute for Earth Science and Disaster Prevention, 3-1 Tennodai, Tsukuba-Shi, Ibaraki-Ken 305-0006 Japan
e-mail: shimada@bosai.go.jp; Tel.: +81-29-863-7622; Fax: +81-29-854-0629

Abstract. The Global Positioning System (GPS) data observed in the Tsukuba 2000 and 2001 GPS/MET dense network campaigns were analyzed. In these campaigns about 75 GPS receivers were installed in a 20 x 20 km area with 1.5km – 3.0km intervals. The 2000 campaign was carried out late fall for 31 days and the 2001 campaign in the summer season for 53 days. GAMIT/GLOBK GPS analyzing software was applied to the analysis. In the analysis of the Tsukuba 2000 campaign data, the time variation of the vertical component changed simultaneously at all the sites, making the correlated motion among the campaign sites. It might be attributed to the fact that constraints on the vertical component of Tsukuba IGS site (TSKB), a fiducial site of the campaign, were too loose. The variation was in the range of 20 – 30 mm with a one-sigma repeatability of 4 – 7 mm. When we examine the vertical component of the baseline vectors of each site with reference to one site, TSKB, the time variation of the vertical components with reference to the TSKB site was much smaller than for single stations. The time variation was in the range of 10 – 20 mm with the one-sigma repeatability of 2 – 6 mm. in this case for the 2000 campaign. The repeatability of the baseline vector among the network sites had a weak dependency on the baseline length, with the zero length repeatability of 3.17 mm and a baseline length dependency of 0.11 ppm. Meanwhile, in the 2001 campaign, the time variation of the baseline height was around 10 mm and almost random. The relationship between the baseline length and vertical repeatability of the baseline vector among the network sites was strong, with the zero length repeatability of 3.36 mm and the baseline dependency of 0.17 ppm during the 2001 campaign.

Keywords: Global Positioning System — Dense-network campaign — Vertical positioning

1. Introduction

In "GPS meteorology project in Japan", we carried out Tsukuba dense network campaigns in 2000 and 2001 using about 75 Global Positioning System (GPS) receivers in 20 x 20 km area with the interval of 1.5km – 3.0km (Shoji *et al.*, 2003a; Iwabuchi *et al.*, 2003; and Shoji *et al.*, 2003b). One of our major objections is (2001) derives the vertical component repeatability as 9.78-mm by analyzing a one-year data set of the GEONET, which is the nationwide GPS network in Japan. Tregoning *et al.* (1999) describes that for GPS measurements in Antarctica for observing postglacial rebound the short-term scatter of the daily vertical solutions is ~7mm at the Beaver Lake stations, using the IGS global network in the same manner as in this paper.

Observations of the vertical component are important to detect not only the magma-related uplift in the volcanic area and postglacial rebound, but also tectonic and earthquake-related vertical motions in the plate boundary zones like the Japanese Islands.

In this study, we analyzed the GPS data, using GAMIT/GLOBK software (King and Bock, 2002; Herring, 2002), and examined the time variations of the resulting site positions and baseline vectors, and the relation between baseline repeatability and baseline length.

2. Data

The Tsukuba 2000 dense network campaign was held for 31 days from October 14th (Day Of Year 287) to November 13th (DOY 318), 2000. Seventy-nine GPS receivers were installed at 75 sites, within an area of 20 x 20 km with a 1.5km – 3.0km interval in and around Tsukuba City, central part of Japan (Fig. 1). The number and type of receivers were: 58 Trimble, 19 Ashtech, and 2 Rogue AOA receivers. The receivers were equipped with following antenna: micro-strip antennas for all the Trimbles and for 2 Ashtechs; choke ring antennas for 17 Ashtechs and all the Rougue AOAs.

The Tsukuba 2001 dense network campaign was conducted for 53 days from July 2nd (DOY 193) to September 2nd (DOY 245), 2001. Seventy-eight receivers were installed at 75 sites similar to those of 2000 campaign. The receivers were: 61 Trimbles, 16 Ashtechs, and 2 Rogue AOAs. All the Trimble receivers used micro-strip antennas, while all the Ashtechs and Rougue AOAs were.using choke ring antennas.

3. Analysis

The GPS data were analyzed using the GAMIT/GLOBK software, following the procedure mentioned in Dong *et al.* (1998). In the GAMIT analysis for the campaign sites ten groups (seven Trimble, two Ashtech, and one Rogue AOA) of stations with similar antenna were formed to cancel out antenna phase errors (Hatanaka *et al.*, 2001). As fiducial sites each group was extended with 7 IGS (International GPS Service) global tracking sites in and around Japan, such as Fairbanks in Alaska (FAIR), Kokee Park in Hawaii

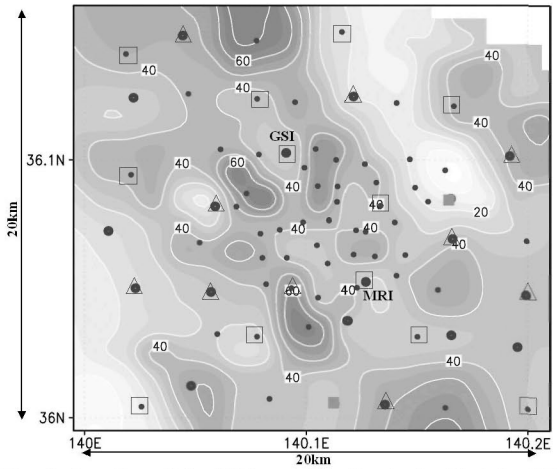


Fig. 1. Location of the GPS receiver sites in the Tsukuba 2000 dense network campaign. Small closed circles – Trimble receiver, large closed circles – Ashtech receiver equipped with Choke ring antenna, closed squares – Ashtech receiver equipped with micro-strip antenna. Open triangle and square – surface meteorological measurements. Contour is topography (unit: meter).

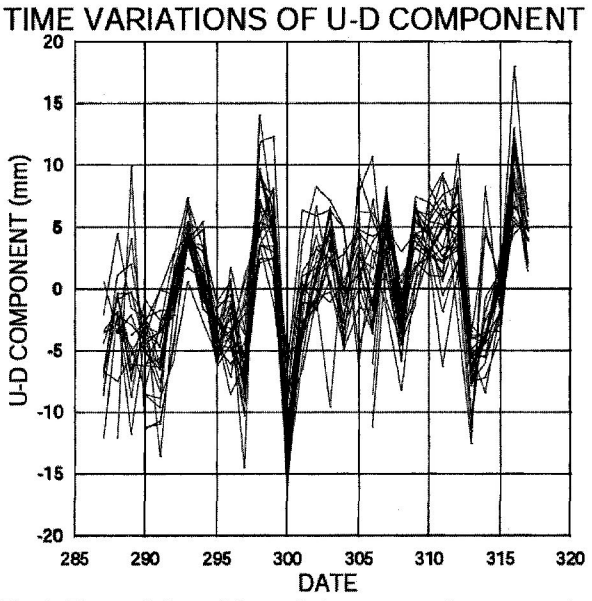


Fig. 2. Time variations of the vertical component of some campaign sites in the Tsukuba 2000 dense network campaign.

(KOKB), Guam, Shanghai, Daejon in South Korea (DAEJ),Usuda in Central Japan (USUD), and Tsukuba in Geographical Survey Institute (TSKB). All the daily data were analyzed on a 24-hours basis. The GLOBK program was used to integrate all local groups with SIO (Scripps Institutions for Oceanography) IGS global network analysis data on a daily basis. In the GLOBK analysis, the horizontal coordination were tightly constrained at 16 global fiducial sites (Onsala in Sweden, Kootwijk in the Netherlands, Wettzell in Germany, Algonquin in Canadian East Coast, Yellowknife in Central Canada, Penticton in Canadian West Coast, FAIR, KOKB, Santiago in Chili, Yarragadee and Canberra in Australia, Guam, Shanghai, DAEJ, USUD, and TSKB), and the vertical coordination loosely constrained because vertical errors are generally very large and systematic. For the satellite ephemeris IGS final solutions were applied.

4. Result

The time variation of the vertical position at the site in the Tsukuba 2000 campaign (Fig. 2) showed a strong correlation because the constraint on the vertical component of TSKB was very loose. The variation was between 20 – 30 mm with a one-sigma repeatability of 4 – 7 mm. Because only TSKB site has the complete tracking data set throughout the campaign, selecting TSKB as the reference site the variation of the vertical component of baseline vectors was analyzed. Fig. 3 shows the time variations of the vertical components of the baseline vectors. We can conclude that the variations in the baselines are smaller and less correlated from for the individual stations. However, some effect common to all baselines can still seen from DOY 293 to DOY 297. This can be attributed to the fact that the Turbo-Rogue eight-channel receiver at reference site (TSKB) was old and its specification was limited. The variation was in the range of 10 – 20 mm with a one-sigma repeatability of 2 – 6 mm (Fig. 4). The sites GM 39 and GM 30 had extraordinary large repeatabilities of 10mm and 8mm respectively. Both of the sites used Ashtech micro-strip antenna instead of choke antenna, therefore the antenna phase center variation did not cancel out. At last, we examined the relationship between the baseline length and repeatability of the baseline vector among the network sites. Fig. 5 gives the repeatability of the vertical component of the baseline vector as function of baseline length, and shows a weak dependency of the repeatability on the baseline length. The dependency is represented as the linear regression formula:

$$3.17mm + 0.11ppm \times L \quad , \qquad (1)$$

where *ppm* is part per million and L the baseline length (mm).

The time variation in the vertical component of the Tsukuba 2001 campaign also showed a strong correlation similarly with the Tsukuba 2000 campaign therefore the figure omitted in this paper. Next, we examined the time variations of the vertical positions of the baseline vectors with respect to a common site. At first, we selected a site located at the north-west end of the network. The time variations showed the much correlated motion, possibly reflecting the network tilt-like motion affected by the correlated network variations. Secondly we selected a site located at the center of the network. Fig. 6 shows the time variations of network sites with respect to GM35, in the center of the 2001 campaign. The variation was around 10 mm in general. However, smaller fluctuations were recorded around DOY 220 while larger fluctuations were seen around DOY 240. We also found that the variations were almost random. It can be due to the homogeneous distribution of the water vapor in the troposphere in the daily-based time scale. We also examined the relationship

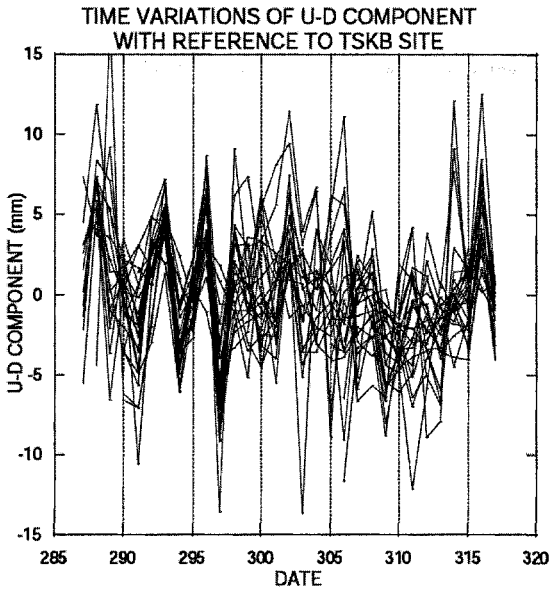


Fig. 3. Time variations of the vertical components of the baseline vectors in the Tsukuba 2000 dense network campaign as in Fig. 2 with the TSKB site fixed.

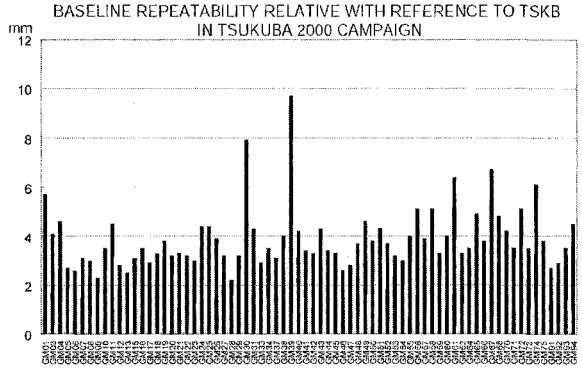


Fig. 4. Vertical repeatabilities of the baseline vectors of the Tsukuba 2000 dense network campaign sites with reference to the TSKB site.

between the baseline length and the repeatability of the baseline vector among the network sites. Fig. 7 shows the baseline length dependency of the repeatability of the vertical component of the baseline vector. A strong dependency with the baseline length was detected. The dependency is represented as the following linear regression formula:

$$3.36mm + 0.17ppm \times L, \qquad (2)$$

where *ppm* is part per million and L the baseline length (mm).

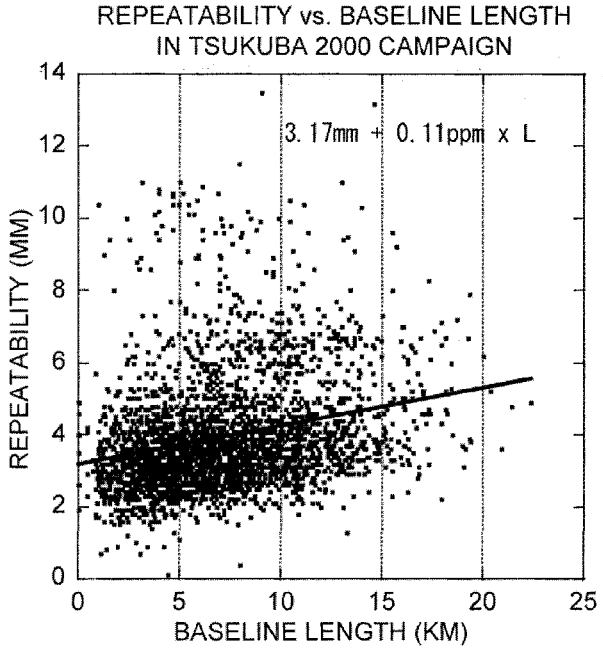


Fig. 5. Baseline length versus repeatability of the vertical component of the all baseline vectors among the Tsukuba 2000 dense network campaign sites.

5. Discussion

Comparing the Tsukuba 2000 and 2001 campaigns, obviously the baseline repeatability in the fall season was better than that of summer season. Because some outliers were detected in the sites where the 2 Ashtechs with different micro-strip antenna were deployed in the 2000 campaign, the baseline repeatability became worse in this analysis. In the summer season, an obvious baseline length dependency existed in the vertical component of the baseline vectors, indicating small scale (1 – 20 km) water vapor turbulence. Using microwave radiometers (MWRs), Aonashi *et al.* (2003) observed the water vapor in this area. From the analysis of the observed data at two sites 10 km apart in the Tsukuba area, they concluded the horizontal scale of the water vapor inhomogeneity was less than 10km. It may be very difficult to model the water vapor inhomogeneity of this scale. One of the targets of both of the campaigns was to investigate the behavior of the vertical repeatability in the range of 1 – 20 km. We found a zero-length repeatability of about 3 mm and a baseline length dependency of 0.1 ppm. Three mm zero-length repeatability was very reasonable value comparing with the about 1 mm of horizontal components repeatability in the recent state of the art, and improved the effectiveness of the antenna grouping approach to cancel out the antenna phase center variations. A zero-length repeatability and baseline length dependency of vertical component would be worse in the mountainous inland Japanese islands than in Tsukuba campaign sites located in a relatively flat terrain.

We need to reduce both the zero-length repeatability and the baseline length dependency. Zero-length repeatability

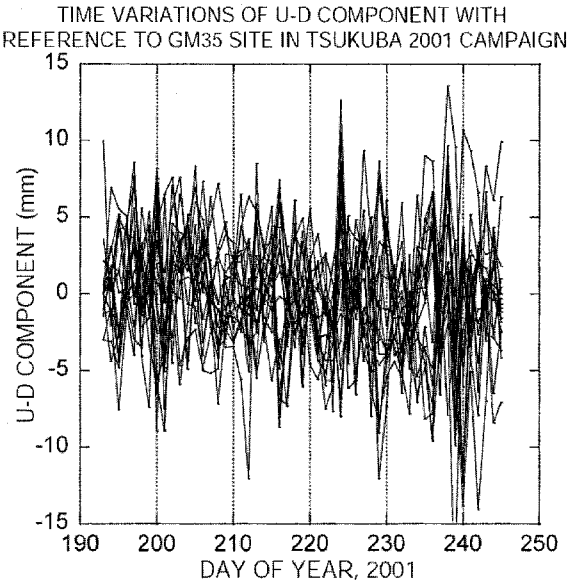


Fig. 6. Time variations of the vertical component with reference to the GM35 site locating at the central area of the network in the Tsukuba 2001 dense network campaign.

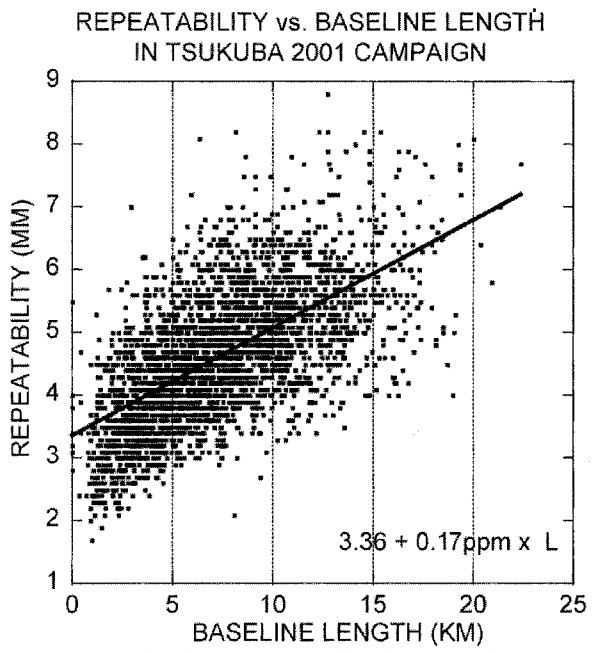


Fig. 7. Baseline length versus repeatability of the vertical component of the all baseline vectors among the Tsukuba 2001 dense network campaign sites.

will be possibly improved by removing multipath by way of hardware, receiver firmware and software. Recent advancement of the receiver firmware with the choke ring antennas may contribute to the reduction of the multipath effect on the precise positioning.

The major problem is the baseline dependency part of the vertical repeatability. The level of 0.1 ppm with 3mm zero-length repeatability obtained in this study is at least one-order worse than the 0.001-0.01 ppm level with 6 mm – 10mm zero-length repeatability commonly obtained in the regional/continental networks (e.g. Hatanaka *et al.*, 2003), which suggests the errors does not increase linearly with distance. Hence, ppm is not a good measure as we divide by baseline length. The water vapor turbulence becomes active 45% or more in the summer than in the fall, according to the comparison between 0.17 ppm in the 2000 campaign and 0.11 ppm in the 2001 campaigns. A dynamic mapping function derived from vertical profiles of the atmospheric numerical weather model may be useful to reduce the dry part of the tropospheric error (Yamagiwa and Hatanaka, 2003) with the baseline range of 30 – 100 km and the time range of seasonal variations. One of the major approaches is to omit the large postfit residuals data with the tight threshold using the azimuth/elevation antenna phase model. Applying the precise mapping function by using the numerical weather prediction through this process should improve the vertical repeatability.

6. Conclusion

GPS data of the Tsukuba 2000 and 2001 dense network campaigns were analyzed. The 2000 campaign was carried out late fall and the 2001 campaign in the summer season.

GAMIT/GLOBK GPS analysis software was applied to the analysis. For GAMIT analysis, IGS sites in and around Japan were used as the fiducial sites. For GLOBK analysis, SIO IGS global network solutions were combined and 16 global IGS sites were employed as the fiducials.

Time variation of the vertical component of the sites in the 2000 campaign showed that the coordinates of the most sites had similar errors, making the correlated motion among the campaign sites because of loose constraint on the vertical component of the TSKB site. The variation was in the range of 20 – 30 mm with a one-sigma repeatability of 4 – 7 mm. The time variation of the vertical components of the campaign site with reference to the TSKB site showed that the time variation reduced the correlated motion. The variation was in the range of 10 – 20 mm with a one-sigma repeatability of 2 – 6 mm. The repeatability of the baseline vector among the network sites had a weak dependency on the baseline length in the 2000 campaign. The dependency was represented 3.17mm for the zero length plus 0.11ppm proportional to the baseline length.

In the Tsukuba 2001 campaign the time variations of the network sites with reference to GM35 located for the central area showed that the variations were around 10 mm in general and almost random. The relationship between the baseline length and vertical repeatability of the baseline vector among the network sites during the 2001 campaign was strong. The zero length repeatability is 3.36 mm and the baseline dependency is 0.17 ppm.

References

Aonashi, K., T. Iwabuchi, Y. Shoji, R. Ohtani, and R. Ichikawa, 2003: Statistical study on precipitable water content variations observed with ground-based microwave radiometers, Proc. Int. Workshop GPS Meteor., Tokyo, Japan.

Iwabuchi, T., Y. Shoji, S. Shimada, H. Nakamura, H. Seko, and K. Aonashi, 2003: Tsukuba GPS dense net campaign observation: Comparison and evaluation of slant wet delays estimated in three types of software, Proc. Int. Workshop GPS Meteor.,Tokyo, Japan.

Dong, D., T. A. Herring, and R. W. King, 1998: Estimating regional deformation from a combination of space and terrestrial geodetic data, *J. Geod.*, **72**, 200-214.

Hatanaka, Y., M. Sawada, A. Horita, M. Kusaka, J. M. Johnson, and C. Rocken, 2001: Calibration of antenna-radome and monument-multipath effect of GEONET – Part 2: Evaluation of the phase map by GEONET data, *Earth Planets Space*, **53**, 23-30.

Hatanaka, Y., T. Iizuka, M. Sawada, A. Yamagiwa, Y. Kikuta, J. M. Johnson, and C. Rocken, 2003: Improvement of the analysis strategy of GEONET, *Bull. Geograph. Survery Inst.*, **49**, 11-37.

Herring, T. A., 2002: GLOBK global Kalman fiter VLBI and GPS analysis program, Mass. Inst. of Technol., Cambridge, Massachesetts.

King, R. W., and Y. Bock, 2002: Documentation for the GAMIT GPS analysis software, Mass. Inst. of Technol., Cambridge, Massachusetts.

Larson, K. M., P. Cervelli, M. Lisowski, A. Miklius, P. Segall, and S. Owen, 2001: Volcano monitoring using the Global Positioning System: Filtering strategies, *J. Geophys. Res.*, **206**, 19453-19464.

Shoji, Y., H. Nakamura, T. Iwabuchi, K. Aonashi, H. Seko, K. Mishima, A. Itagaki, R., Ichikawa, R. Ohtani, 2003a: Tsukuba GPS dense net campaign observation: Improvement of GPS analysis of slant path delay by stacking one-way postfit phase residuals, Proc. Int. Workshop GPS Meteor., Tokyo, Japan.

Shoji, Y., K. Aonashi, H. Nakamura, R. Ohtani, T. Iwabuchi, R. Ichikawa, 2003b: Tsukuba GPS dense net campaign observation: analysis of slant water vapor by GPS and comparison with water vapor radiometer, Proc. Int. Workshop GPS Meteor., in Tokyo, Japan.

Tregoning, P., A. Welsh, H. McQueen, and K. Lambeck, 2000: The search for postglacial rebound near the Lambert Glacier, Antarctica, *Earth, Planets Space*, **52**, 1037-1041.

Yamagiwa, A., and Y. Hatanaka, 2003: Mapping functions generated from numerical model and their impact on seasonal variations in GPS analysis, Proc. Int. Workshop GPS Meteor., Tokyo, Japan.

Horizontal Crustal Motion Estimates for the UK inferred from GPS: Preliminary Results

A.I. Waugh, F.N. Teferle, R.M. Bingley, and A.H. Dodson

Institute of Engineering Surveying and Space Geodesy, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

Abstract. This paper presents preliminary estimates of horizontal crustal motion inferred from a GPS network in the UK. The British Isles are on the western part of the Eurasian plate and the effects of active tectonism are expected to be small. However, the British Isles are believed to undergo non-rigid plate behaviour due to glacial isostatic adjustment (GIA) occurring in northern Britain. Non-rigid plate behaviour can generally be shown to cause velocity biases when compared to velocities derived from models such as NNR-NUVEL-1A, which are based upon assumed rigid plate behaviour. As a first assessment, in this paper, horizontal station velocities obtained from the GPS analysis are compared to the ITRF2000 angular velocity of rigid Eurasia.

Keywords. British Isles, Horizontal Crustal Motions, GPS, Eurasian Plate

Introduction

It is thought that, in common with most of Western Europe, the British Isles behave in a stable manner, with active tectonism being negligible and GIA being confined to northern Britain. At present the degree to which the GIA process induces additional horizontal motions in northern Britain has not been determined. However it is widely accepted that the horizontal GIA component will be a larger signal than the vertical. It should, therefore, prove easier to detect and measure, and be useful for constraining GIA models.

The establishment of GPS stations, both continuous (CGPS) and episodic (EGPS), around the UK has allowed a quantitative analysis of horizontal crustal motion in the British Isles to be made. In this study, data from varying temporal scales has been used in producing estimates of horizontal crustal motion. For a first assessment the displacement vectors obtained from GPS are compared to predictions based upon the ITRF2000 angular velocity for the Eurasian plate (Altamimi et al., 2002; Altamimi et al., 2003). As expected there is good agreement between the predicted and the observed displacements, reinforcing the view that the British Isles are situated on a stable part of the Eurasian plate. Discrepancies, where they do occur, could be a result of GPS measurement errors, reference frame errors, biases within the ITRF2000 angular velocity or actual differences between the crustal motion and the underlying plate movement.

The GPS Network

In carrying out this study the following UK GPS stations were used for analysis:

- 16 CGPS stations observed since 1997-1998, including 6 at tide gauge sites (CGPS@TG) and 10 at inland sites.
- 11 EGPS stations at tide gauge sites (EGPS@TG) observed from 1991 to 2003.

The network provides a good spatial coverage over the UK and should, therefore, provide a good description of horizontal crustal motion in the British Isles.

GPS Processing Strategy

All GPS processing was carried out using the Institute of Engineering Surveying and Space Geodesy (IESSG) GPS Analysis Software version 2.4 (Stewart et al, 2002). In recent years the IESSG has invested a great deal in automating its software to process the data from the ever increasing UK GPS network. The data from the 27 GPS stations were processed as a series of sub-networks, depending on data availability. Without exception, data from the IGS stations at Onsala, Kootwijk, Wettzell and Villafranca were included in each processing sub-network. To process the UK GPS data at the GPS estimation stage the IGS stations were motioned to the correct observation epoch but were not constrained. In this way a "loosely" constrained network is formed through the fixing of the IGS precise ephemeris. At the post processing stage the ITRF2000 coordinates of the IGS stations were motioned to the observation epoch by applying the appropriate velocity and these were then constrained in order that the vectors from the

GPS estimation stage could be used to compute coordinates for the UK stations in ITRF2000.

In geophysical monitoring where millimetre accuracy is required, it is important to mitigate against all error sources, For this reason, the final International GPS Service (IGS) precise ephemeris was used along with the L1/L2 ionospherically free observable. Furthermore, solid Earth tides and ocean tide loading were modelled, according to IERS standards, and tropospheric delay was mitigated by using Saastamoinen model as an *a-priori* estimate and then solving for a correction to this using a Kalman filter approach.

Coordinate Time Series Analysis

The daily ITRF2000 coordinate estimates output from the GPS processing was used to form coordinate time series. In order to obtain estimates of horizontal station velocities for the EGPS stations within the network, a weighted linear regression was employed with the accompanying velocity uncertainties determined using equation 33 in Dixon (1991).

To take account of the presence of coloured noise in CGPS coordinate time series, the technique used to determine horizontal station velocities for CGPS stations was different. In this case, a Maximum Likelihood Estimator (MLE) was used, with simultaneous estimation of a linear trend, annual signal, coordinate offset magnitudes and the amplitudes of white noise (WN) and power-law noise (PLN) (Williams 2003).

Results

Estimates of horizontal crustal motions for UK CGPS stations, based on the horizontal station velocities computed in ITRF2000 using MLE with a WN and PLN model, are presented in *Table 1* together with their 1σ uncertainties. GPS velocity measurements are determined as absolute and for this reason these first assessments are made against absolute modelled plate velocities. For comparison, the ITRF2000 angular velocity for the rigid rotation of Eurasia (Altamimi et al., 2003) has been chosen and denoted as predicted velocities in *Table 1* and *Figure1*. Until the release of the ITRF2000 angular velocities the plate motion model to compare GPS results against was the NNR-NUVEL-1A, (adapted from NNR-NUVEL-1 (Argus and Gordon, 1991), according to DeMets et al., (1994)), as recommended by International Earth Rotation Service (IERS) standards, (McCarthy, 1996). Although, the ITRF2000 angular velocities are globally aligned to NNR-NUVEL-1A in terms of orientation time evolution, disagreement between the models may exceed 3 mm/yr for regional site velocities (Altamimi et al., 2003). With this in mind, the authors compared their observed station velocities to the predicted Eurasian ITRF2000 angular velocity. As a first step, the analysis of the UK CGPS results revealed that these have a mean

Table 1: Predicted and observed station velocities. Predicted velocities stem from the ITRF2000 angular velocity of the Eurasian plate (Altamimi et al., 2003). Observed velocities have been computed for the ITRF2000 coordinate time series using MLE with a WN and PLN model (Teferle, 2003). All velocities are in mm/yr and bearings are in degrees.

CGPS	Predicted station velocities.				Observed station velocities			
Station	dN	dE	Velocity	Bearing	dN	dE	Velocity	Bearing
ABER	15.1	15.0	21.3	44.5	15.0 +/- 0.2	16.7 +/- 0.4	22.4	48.0
ABYW	15.2	16.1	22.1	46.6	15.9 +/- 0.2	17.9 +/- 0.4	23.9	48.4
BARK	15.0	17.2	22.8	48.8	15.5 +/- 0.4	18.4 +/- 0.2	24.0	50.0
CAMB	15.2	16.6	22.5	47.4	15.9 +/- 0.2	17.3 +/- 0.5	23.5	47.7
DUNK	15.1	16.8	22.6	47.8	16.4 +/- 0.7	17.8 +/- 1.0	24.2	47.3
HEMS	14.9	17.3	22.8	49.0	14.5 +/- 0.5	19.4 +/- 0.5	24.2	53.2
HERS	15.1	17.5	23.1	49.1	12.9 +/- 0.6	19.3 +/- 0.8	23.2	56.3
IESG	15.1	16.6	22.4	47.5	14.3 +/- 0.1	17.9 +/- 0.4	22.9	51.3
LIVE	15.1	16.0	22.0	46.5	13.8 +/- 0.7	15.6 +/- 0.9	20.8	48.5
LERW	15.1	14.1	20.7	43.1	16.4 +/- 0.3	13.6 +/- 0.6	21.3	39.6
LOWE	15.0	17.3	22.8	49.0	15.0 +/- 0.1	18.9 +/- 0.8	24.1	51.6
NEWL	15.2	16.6	22.5	47.4	16.3 +/- 0.3	17.1 +/- 0.6	23.6	46.4
MORP	15.1	15.7	21.8	46.0	17.0 +/- 0.5	14.6 +/- 0.7	22.4	40.8
NSTG	15.1	15.8	21.8	46.2	14.3 +/- 0.3	19.5 +/- 0.5	24.1	53.7
SHEE	15.0	17.5	23.0	49.2	15.2 +/- 0.2	18.6 +/- 0.2	24.0	50.8
SUNB	15.1	17.2	22.9	48.6	15.4 +/- 0.3	18.0 +/- 0.4	23.7	49.5
Mean	15.1	16.4	22.3	47.3	15.2	17.5	23.3	48.8
Stdev	0.1	1.0	0.7	1.8	1.1	1.7	1.1	4.3

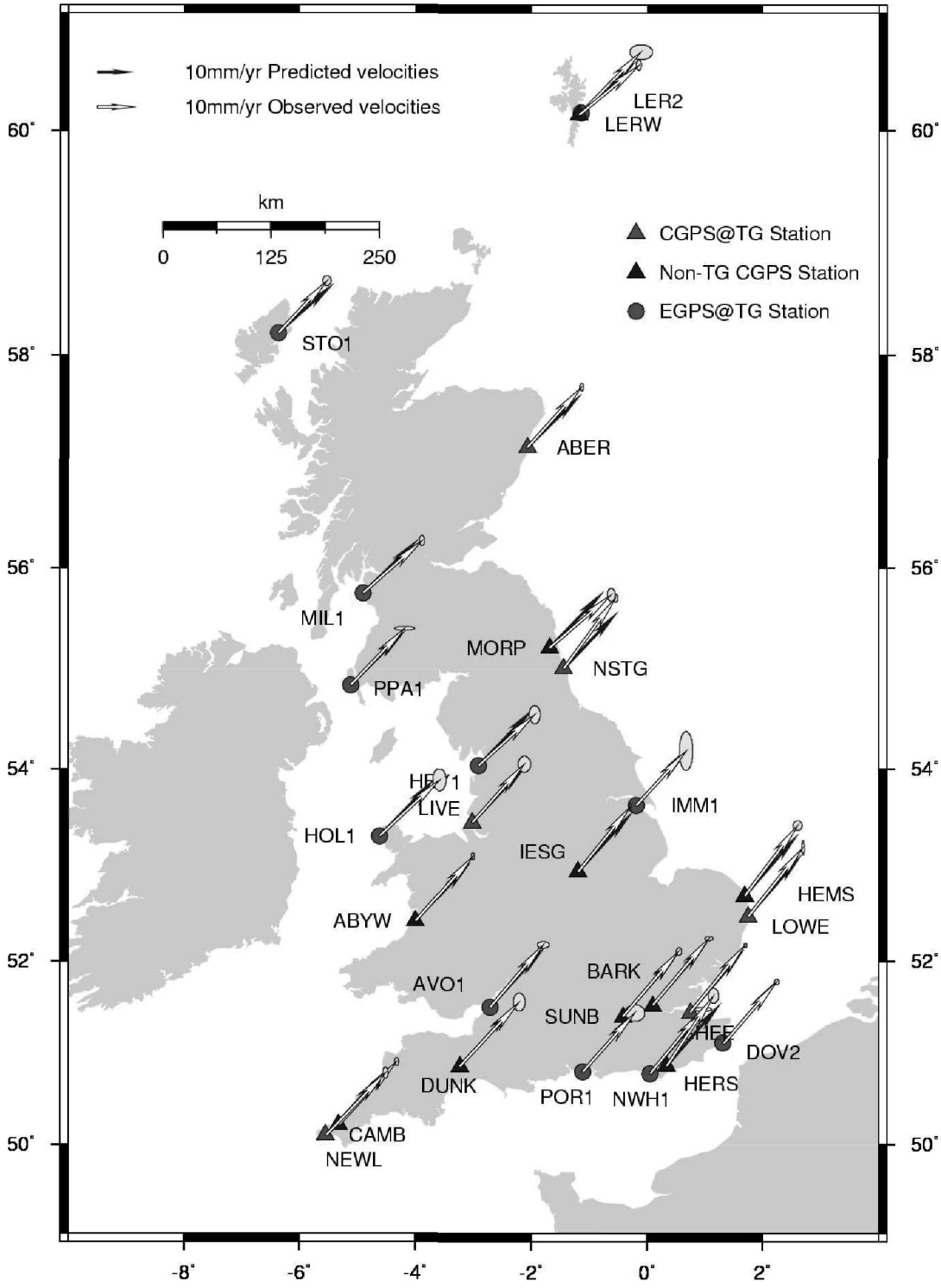


Figure 1: Plot of predicted ITRF2000 observed GPS ITRF2000 velocities for UK GPS stations, for both CGPS and EGPS stations. Predicted velocities are those computed using the ITRF2000 angular velocity of the Eurasian plate. Observed velocities are GPS estimates of station velocity computed in ITRF2000.

Table 2: Predicted and observed station velocities for UK EGPS stations. Predicted velocities stem from the ITRF2000 angular velocity (Altamimi et al., 2003). Observed velocities have been computed from the ITRF2000 EGPS campaign solution. All velocities are in mm/yr and bearings are in degrees.

EGPS	Predicted ITRF2000 angular velocity of Eurasian plate				Observed GPS ITRF2000 velocities			
Station	dN	dE	Rate	Bearing	dN	dE	Rate	Bearing
AVO1	15.2	16.7	21.3	47.6	14.6 +/-0.3	17.1 +/- 0.7	22.5	49.6
DOV2	15.0	17.7	22.1	49.5	14.5 +/- 0.1	16.9 +/- 0.2	22.3	49.3
HEY1	15.2	15.8	22.8	46.0	15.4 +/- 0.3	14.0 +/- 0.7	20.8	42.3
HOL1	15.2	15.7	22.5	45.8	16.4 +/- 0.3	14.9 +/- 0.5	22.1	42.2
IMM1	15.1	16.6	22.6	47.6	13.7 +/- 0.2	14.9 +/- 0.6	20.2	47.5
LER2	15.1	14.1	22.8	43.1	16.4 +/- 0.3	16.7 +/- 0.3	23.3	45.6
MIL1	15.2	14.8	23.1	44.1	16.2 +/- 0.4	14.3 +/- 0.8	21.6	41.4
NWH1	15.1	17.5	22.4	49.2	14.5 +/- 0.3	17.7 +/- 0.6	22.9	50.7
POR1	15.1	17.3	22.0	48.7	14.3 +/- 0.3	16.2 +/- 0.5	21.6	48.6
PPA1	15.1	15.1	20.7	44.7	14.8 +/- 0.2	15.6 +/- 0.5	21.5	46.5
STO1	15.2	13.6	22.8	41.7	13.3 +/- 0.2	14.3 +/- 0.4	19.5	47.1
Mean	15.1	15.9	22.3	46.2	14.7 +/- 0.3	15.8 +/- 0.6	21.6	47.0
Stdev	0.1	1.4	0.7	2.6	1.1	1.3	1.7	3.0

rate of movement of 23.3 mm/yr on a bearing of approximately 49 degrees.

However, on the whole there is good agreement between the CGPS results and the geodetically derived model, which will be considered in more detail in the appropriate section. Those stations that show anomalous results i.e. differing significantly from the ITRF2000 angular velocity for the Eurasian plate, will be investigated in future work.

One such case where there is significant disagreement between the predicted and observed velocities is at NSTG. At this station the differences between observed and predicted velocities are 0.8 and –3.7 mm/yr for the North and East components, respectively. The possible reasons for this relate to the physical environment that the CGPS receiver inhabits. NSTG is situated in a harbour area and is used for geophysical monitoring of vertical crustal motions. This station has been shown to be severely effected by radio frequency interference and multipath (Teferle et al., 2003).

Estimates of horizontal crustal motions for UK EGPS stations that have been observed since 1991 are presented in *Table 2*, again based on horizontal station velocities computed in ITRF2000. As can be seen from this table, for the EGPS stations there is an average rate of motion of 21.6 mm/yr on a bearing of about 47 degrees.

On the whole, these values compare well with the results for the CGPS stations given in *Table 1*. Differences, where they occur, could be the result of factors associated with EGPS campaigns. Principally, in the case of EGPS campaigns, these mainly take the form of human induced errors.

Discussion

Figure 1 shows the first estimates of horizontal crustal motion for the British Isles, inferred from GPS.

From these results it is possible to conclude that the rate and direction of movement of both CGPS and EGPS stations is very strongly latitude dependent, with both velocity and bearing increasing towards the south of the British Isles. This is in line with the movement expected as the angular distance increases away from the Euler pole of the Eurasian plate.

In a further test the horizontal crustal motions were computed relative to the processing reference frame defined by Kootwijk, Onsala, Wettzell and Villafranca. This is carried out by subtracting the average rigid plate motion of these four IGS stations (as calculated using the ITRF2000 angular velocity) from the predicted velocities of the GPS stations. Also calculated for the sake of comparison was the difference of the observed station velocities and the average motion of the four IGS stations computed using their published ITRF2000 velocities. The results of these two relative tests can be viewed in *Figure 2*. A consistent westwards motion in the range of 0.5 to 5.8 mm/yr is seen for the predicted relative station velocities. However, the observed relative station velocities show large variations. The reason for plotting the relative motion of the UK GPS stations against the processing reference frame was to investigate whether any systematic effects related to the reference frame definition had propagated into the observed station velocities. From the current

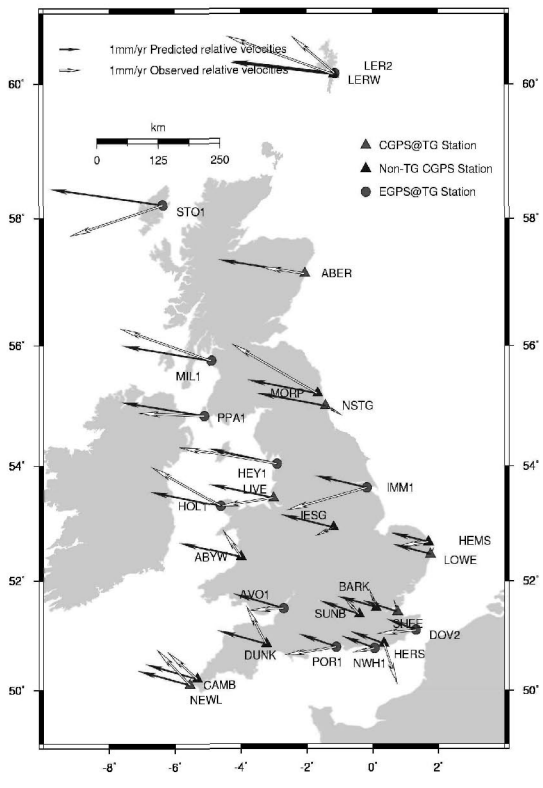


Figure 2: Predicted and observed relative ITRF2000 velocities for UK GPS stations.

agreement in Figure 2, no conclusions with respect to these biases can be made and further investigations are required.

Conclusions and Further Work

In conclusion these preliminary results presented in this paper reinforce the view that the British Isles are situated on a stable part of the Eurasian plate. Further studies will be carried out as coordinate time series are extended and more CGPS stations become available in the UK. This should enable a more rigorous assessment of how suitable the ITRF2000 angular velocity, that describes the rigid motion of Eurasia, is for describing the motion of the British Isles. In recent years, a number of authors have questioned the degree to which Western Eurasia departs from true rigid lithosphere behaviour. In their analysis of Europe, Marotta and Sabadini (2002) concluded that both GIA and active tectonism are presently leading to a complex spatial distribution of extensive and compressive forces of the Western Eurasian lithosphere. Active deformation of the continental lithosphere is thought to be occurring in northern Britain due to GIA, however these current results do not conclusively reveal an underlying GIA signal in the horizontal crustal motion.

At present there are six GPS stations in the area thought to be undergoing isostatic adjustment, but only two of these are CGPS stations. However, in partnership with the Proudman Oceanographic Laboratory (POL) the IESSG plans to establish seven further CGPS stations in Scotland over the next 12 months, four of which will be at tide gauges and three of which will be at specific inland sites in the area of maximum uplift. It is hoped that these stations will allow the authors to accurately determine the degree to which northern Britain departs from rigid lithosphere behaviour.

Acknowledgements

The EGPS campaigns, and CGPS stations at tide gauges, were funded by the UK Department of Environment, Food and Rural Affairs (DEFRA) through their long term commission with POL. Data from UK CGPS stations not funded by DEFRA were made available through the British Isles GPS archive Facility (BIGF) which is funded through the Natural Environment Research Council (NERC).

The authors would like to express their appreciation to all members of the IESSG, past and present, who since 1991 have given up many hours of their own time to conduct the episodic GPS campaigns. Furthermore, thanks should go to Simon Williams for kindly providing the MLE software to the authors.

Thanks also to an anonymous reviewer for useful suggestions to improve the content of this article.

References

Altamimi, Z., P. Sillard, and C. Boucher, ITRF2000: A new release of the International Terrestrial Reference Frame for earth science applications, J. Geophys. Res., 107(B10), 2214, doi: 10.1029/2001JB000561, 2002.

Altamimi, Z., P. Sillard, and C. Boucher, The impact of a No-Net-Rotation Condition on ITRF2000, *Geophys. Res. Lett.*, 30, 1064-1068, doi: 10.1029/2002GL016279, 2003.

Argus, D. F., and R. G. Gordon, No-Net-Rotation Model of Current plate Velocities Incorporating Plate Motion Model Nuvel-1, *Geophys. Res. Lett., 18*, 2038-2042, 1991.

DeMets, C., R. G. Gordon, D F. Argus, and S. Stein, Effect of recent revisions to the geomagnetic reversed time scale on estimates of current plate motions, *Geophys Res, Lett*, 21, pp. 2191-2194, 1994.

Dixon, T. H., An introduction to the Global Positioning System and some Geological Applications, *Reviews of Geophysics*, 29 (2), 249-276, 1991.

Marotta, A. M., and R. Sabadini, Tectonic versus glacial isostatic deformation in Europe, *Geophys. Res. Lett., 29*, 2002.

McCarthy, D., (Ed) *IERS Conventions*, *IERS Tech. Note 21*, Obs. de Paris, Paris 1996.

Stewart, M P, Ffoulkes-Jones, G H, Chen, W, Ochieng, W Y, Shardlow, P J, Penna, N T, and Bingley, R M, GPS Analysis Software (GAS) Version 2.4 User Manual, *IESSG Publication,* University of Nottingham, UK, 2002.

Teferle, F.N., Strategies for Long-term Monitoring of Tide Gauges using GPS, *PhD Thesis*, University of Nottingham (2003).

Teferle, F.N., Bingley, R.M., Dodson, A.H., Apostilidis, P., and Staton, G., RF Interference and Multipath Effects at Continuous GPS Installations for Long – Term Monitoring of Tide Gauges in UK Harbours. (ION).

Wessel, P., and W. H. F. Smith, New version of generic mapping tools released, *Eos Trans, AGU,* 76, 329, 1995.

Williams, S. D. P., The Effect of Coloured Noise on the Uncertainties of Rates Estimated From Geodetic Time Series, *Journal of Geodesy*, 76 (9-10), 483-494.

Results of the SIRGAS Campaign 2000 and Coordinates Variations with Respect to the 1995 South American Geocentric Reference Frame

H. Drewes, K. Kaniuth
Deutsches Geodätisches Forschungsinstitut (DGFI), Marstallplatz 8, 80539 München, Germany
e-mail: drewes@dgfi.badw.de

C. Völksen
Bayerische Kommission für die Internationale Erdmessung (BEK), Marstallplatz 8, 80539 München, Germany

S.M. Alves Costa, L.P. Souto Fortes
Instituto Brasileiro de Geografia e Estatística (IBGE), Av. Brasil 15671, Rio de Janeiro, Brazil, 21241-051

Abstract. The geocentric reference system for the Americas (SIRGAS) was initially realised for South America only. A GPS campaign in 1995 covered 58 sites. In 2000 these stations were re-observed and the network was extended to North and Central America as well as the Caribbean. The objectives of the project were completed by the establishment of a unified vertical reference system connecting the classical national height systems. The processing of the 2000 observation data was performed by three analysis centres at DGFI and BEK in Munich as well as IBGE in Rio de Janeiro. The Bernese and GIPSY/OASIS software packages were used. The processing and the comparisons with the 1995 coordinates are described.

Keywords: Terrestrial reference frame, GPS, South America, unified height system

1 Introduction

The South American geocentric reference frame (Sistema de Referencia Geocéntrico para América del Sur, SIRGAS) was installed by a ten-day GPS campaign in May 1995 (SIRGAS 1997). A total of 58 stations were observed. The coordinates were obtained in the ITRF94 at epoch 1995.4 with a precision of ±3 ... ±6 mm. In May 2000 the first repetition campaign was performed extending the project to North and Central America as well as to the Caribbean (Sistema de Referencia Geocéntrico para las Americas, Luz et al. 2002). The geodetic objectives were extended to the establishment of a unified vertical reference frame connecting all the national height systems (Drewes et al. 2002). A total of 184 stations were observed during ten days including all the tide gauges defining a national height system and some levelling points at the borders between neighbouring countries (figure 1).

2 Data Processing and Performance

The data processing of the entire SIRGAS 2000 network was done by three analysis centres: DGFI, BEK and IBGE. As a large variety of receiver and antenna models were involved, considerable efforts were necessary to identify the correct receivers and antenna types in order to make the RINEX data files compatible with the IGS naming conventions. The same holds for the antenna heights where the type of measurement (slant or vertical) and the reference points were not as well documented as in the permanent GPS networks. Both tasks were done in collaboration between DGFI and IBGE. First results of the data processing were reported during an International Symposium (IAG) in Cartagena, Colombia (Costa et al. 2002, Kaniuth et al. 2002).

For the final solution DGFI and IBGE used the latest version of the Bernese software. The principal characteristic of this system is the processing of phase differences between stations and satellites, i.e., the double difference approach (Hugentobler et al. 2001). The third analysis centre, BEK, used the GIPSY/OASIS II software developed at Jet Propulsion Laboratory (JPL) which processes undifferenced code and phase observations (Webb and Zumberge 1997). The precise point positioning strategy (Zumberge et al. 1997) was applied. Some processing features were aligned to each other, but in general each analysis centre was free to select certain options and settings according to its experience. The common features include:

- Application of a 10° elevation angle cut-off in all network adjustments and the tropospheric mapping function of Niell (1996).

- Application of the relative antenna phase centre offset and phase centre variation models proposed by IGS, if necessary supplemented by calibrations performed at the National Geodetic Survey (NGS), USA (http://www.igscb.jpl.nasa.gov/igscb/station/general/igs_01.pcv).

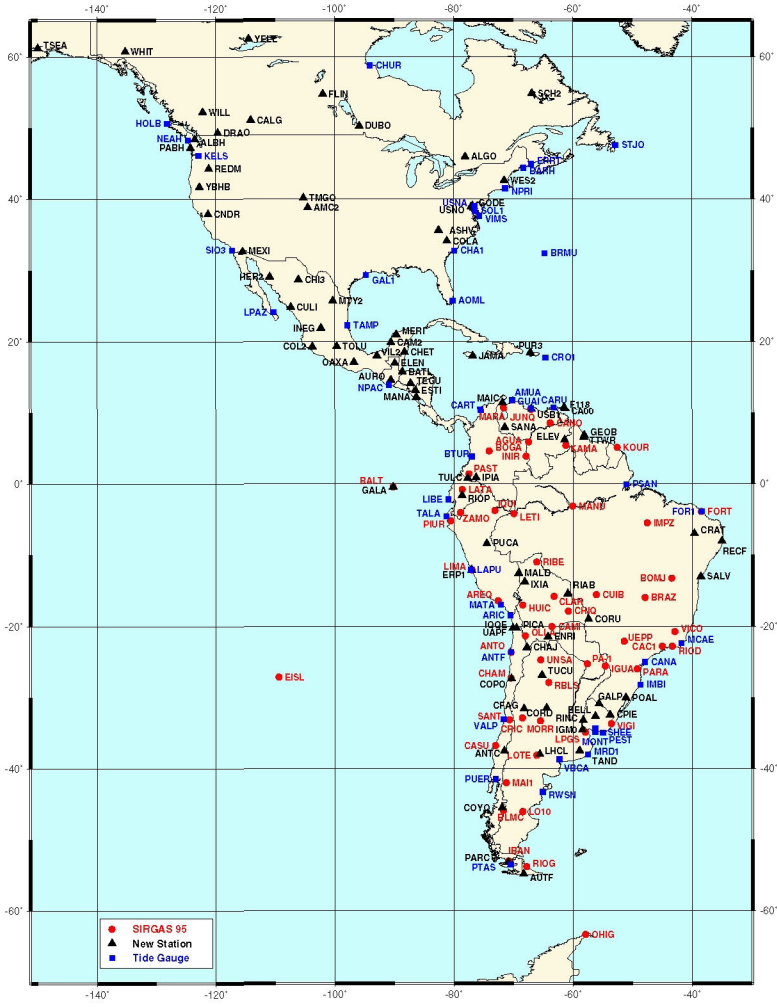


Fig. 1: SIRGAS stations 1995 and 2000

Regarding satellite orbits, satellite clock offsets with respect to GPS time and Earth orientation parameters, both DGFI and IBGE referred the network adjustments to the combined IGS solutions. BEK used the corresponding JPL products including a subsequent empirical transformation to the International Terrestrial Reference Frame 2000 (ITRF2000) also provided by JPL. In case of the Bernese software the receiver clock errors are estimated with sufficient accuracy in a code single point positioning adjustment, whereas in GIPSY/OASIS II they are estimated together with all the other parameters in the precise point positioning solutions.

The different algorithms used in the two software packages, least squares adjustment and square root information filter, respectively, lead to different approaches for modelling the tropospheric zenith delay. In the GIPSY/OASIS II solution the variation is modelled as a random walk process constrained to 10 mm/$\sqrt{}$ hour. The Bernese software models the zenith delay as a step function allowing a smoothing of the variation with time by setting certain parameters appropriately. Both DGFI and IBGE set the time windows to two hours and left the variation practically totally free because the available satellite constellation allows an unconstrained estimation. Major processing

differences between the two Bernese solutions relate to the observation weighting and the outlier handling:

- According to previous experience (Kaniuth et al. 2002), DGFI did not down-weight the observations with decreasing elevation angle and rejected data only after editing the baseline adjustment residuals;
- IBGE weighted the data according to cosine zenith distance and applied the outlier rejection strategy available in the Bernese software.

GIPSY/OASIS II also does not apply elevation dependent weighting but all phase observations with residuals exceeding 2.5 cm are rejected.

The daily network adjustments using the double differencing technique could not be performed in one step. Both analysis centres had to partition the network, IBGE did it in nine and DGFI in three blocks, which were then combined on the normal equation level. The only small deficiencies of this strategy are:

- Tropospheric zenith delays at the sub-network junction points are set up twice and independently.
- The correlation due to including observations of junction points into two baselines is neglected.

As examples, table 1 and figure 2 display the point positioning performance achieved in terms of daily repeatability with respect to the individual ten-day solution for four South American stations which were already included in the SIRGAS 1995 campaign. The sites Bogotá (BOGA) and Latacunga (LATA) are situated in the equatorial region where the ionospheric disturbances were rather high during the 2000 observation epoch. The other two (HUIC and MAII) are in southern latitudes.

In the horizontal components there are hardly any differences between the two Bernese solutions with an average daily repeatability slightly better than ±3 mm. In the height component the differences are also small. GIPSY/OASIS II shows the best repeatability of all three solutions in the north component where the network extends to more than 13.000 km. The inferior performance in the east component is easily explained by the fact that in the precise point positioning mode no phase ambiguities can be resolved.

3 Datum Realisation and Combination of the Solutions

The solutions of the three analysis centres have to be referred to a unique datum. Therefore it was necessary to select an adequate set of stations to be used for the reference frame realisation. The criteria for selecting these fiducial sites were:

- Good tracking performance during the entire campaign in order to propagate the fiducial position accuracy fully into the SIRGAS network.
- Co-location sites with VLBI and/or SLR are preferred, assuming that different space techniques contribute in particular to accuracy and reliability of the vertical position and velocity components.
- The position and velocity estimates of the ITRF2000 at co-location sites should rely on accurate local ties between the GPS, SLR and VLBI reference points.
- Highest accuracy level in ITRF2000, based on sufficiently long time series of the observations and redundant solutions, e.g. vertical velocity standard deviation ≤ 1 mm/a.
- The distribution of the fiducials should be homogeneous over the entire network area.

These criteria lead to the selection of 14 fiducial sites. They are displayed in figure 3. Each analysis centre realised the reference frame individually. The Bernese type analysis centres extended their normal equation system by seven Helmert transformation parameters. The estimation of these parameters is based on the 14 fiducial sites which realise the datum of the Bernese solutions in the ITRF2000 at the observation epoch.

Table 1: R.M.S. repeatability in North (N), East (E), and Height (H) of daily solutions with respect to the entire ten-day adjustment [mm]

Station ID	Location		BEK			DGFI			IBGE		
	φ [°]	λ [°]	N	E	H	N	E	H	N	E	H
BOGA	4.64	74.08	1.4	5.0	6.6	2.5	3.0	4.6	3.2	2.8	5.7
LATA	-0.82	78.62	1.9	5.0	4.0	2.9	2.5	4.9	2.9	2.0	7.2
HUIC	-17.03	68.45	3.8	6.8	3.9	2.3	2.5	4.7	2.5	2.7	4.9
MAII	-42.02	71.20	2.5	5.5	6.7	3.9	2.3	5.0	4.1	2.9	5.4

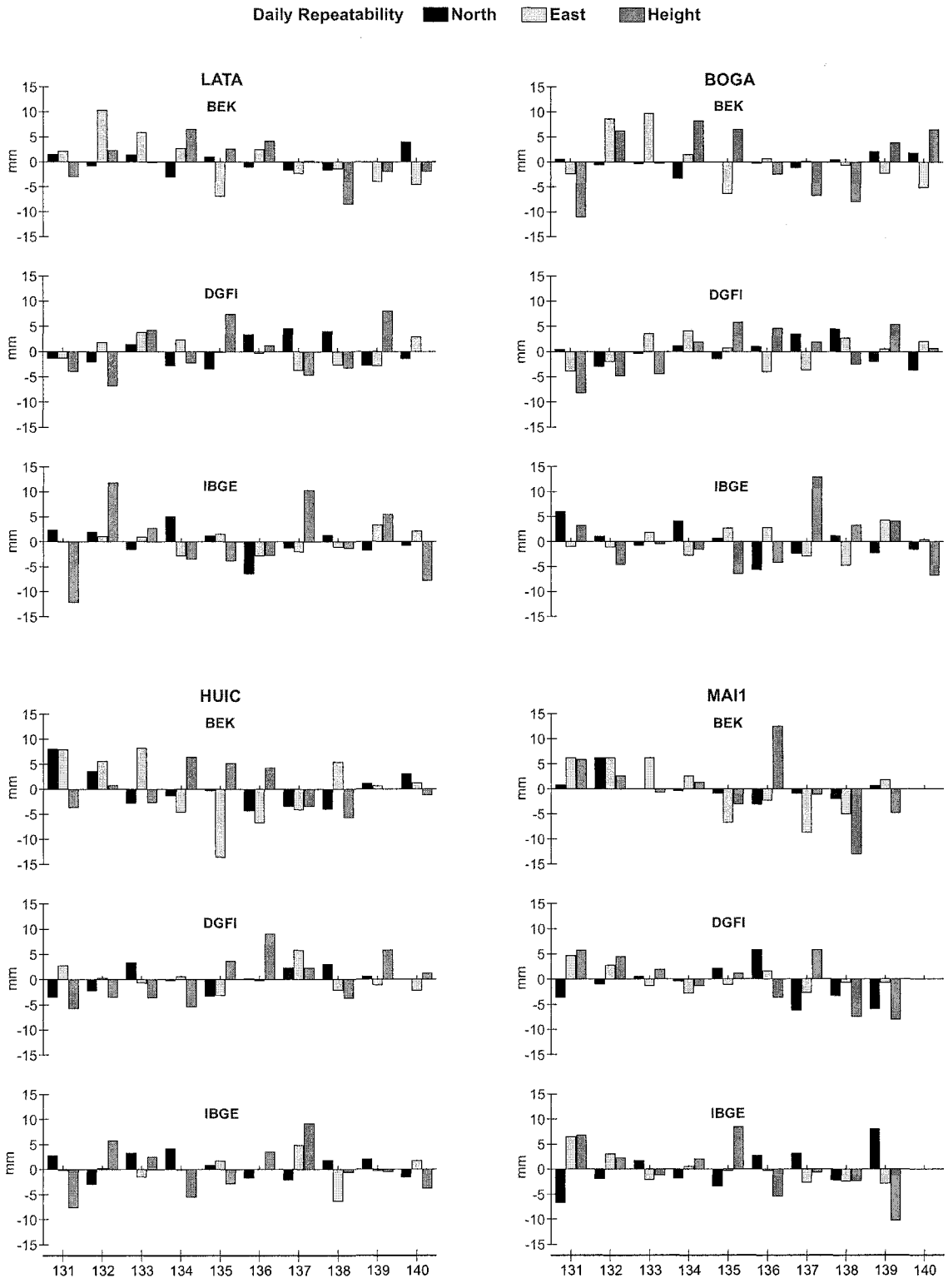


Fig. 2: Daily repeatability in four selected SIRGAS 2000 stations

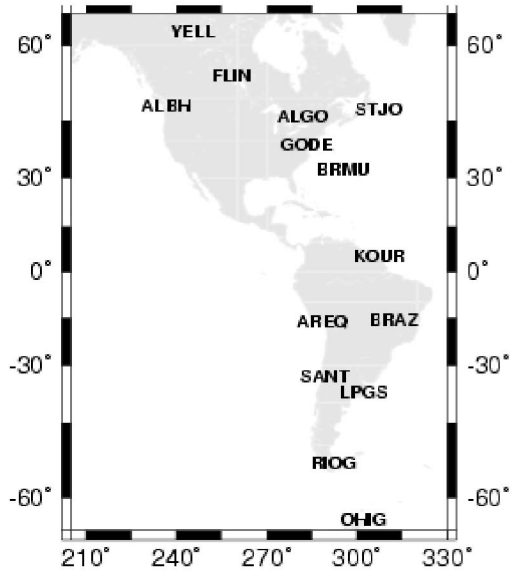


Fig. 3: Sites selected for the reference frame realisation. The 4-character IDs are those used by the IGS.

As mentioned earlier, the GIPSY/OASIS II solution was processed in a non-fiducial system realised by the JPL orbits and satellite clocks. Usually a transformation into a fiducial frame is carried out using parameters supplied by JPL. In our case the non-fiducial solution of each individual day has been transformed to the 14 fiducial sites to realise the ITRF2000 at the observation epoch. Following this procedure any possible distortion of the network is minimised.

The next step is the combined adjustment of the solutions of the three SIRGAS analysis centres. Usually Bernese type solutions are realised by stacking the normal equations of the individual sessions and then solving for the parameters. This very flexible approach could not be used because GIPSY/OASIS II does not deliver normal equations since it is based on a filter.

As a consequence the combination was done on the basis of coordinates and their covariance matrices. This is not a disadvantage because the adjustment on the level of normal equations containing only coordinates and the adjustment based on coordinates and their covariances should yield identical results. Therefore the individual analysis centres generated coordinates and covariance information in SINEX format. The weighting of the individual solutions prior to the combination remains as the major problem.

GIPSY/OASIS II estimates its statistics on the level of coordinates misclosures from the individual session solutions and from some predefined settings (e.g. phase and code noise). The observation interval was set to 300 seconds only while the Bernese type analyse centres used a much higher rate of 30 seconds, which also has an impact on the error estimates of the coordinates. The Bernese statistics strongly depend on the degree of freedom stemming from the large amount of phase observation and the stochastic model assuming uncorrelated observations. Therefore they are usually by far too optimistic; this also holds for the GIPSY/OASIS II coordinates estimates.

Therefore a re-scaling of the individual variance/ covariance matrices was applied. The scale factor is derived from the mean point errors of the individual solutions. This is based on the assumption that the adjusted "absolute" coordinates should have the same level of accuracy in all the individual solutions. It seems to be the most reasonable approach which treats the contribution of each analysis centre evenly. As a result the solutions of the Bernese type were scaled down by factors of 5.20 (IBGE) and 5.65 (DGFI) with respect to the GIPSY/OASIS II solution. The combined adjustment of these sets provides the final set of coordinates.

In order to evaluate the combined solution a comparison with the individual solutions is done by performing 7-parameter Helmert transformations. Table 2 gives the r.m.s. deviation of the north, east and height components with respect to the combined solution. All three individual solutions appear to be represented well in the combination.

Table 2: R.M.S. agreement in north, east and height components between the individual solutions and the combined solution after applying a 7-parameter Helmert transformation [mm].

	North	*East*	*Height*
BEK	± 2.2	± 4.0	± 6.8
DGFI	± 2.6	± 3.7	± 7.0
IBGE	± 2.6	± 3.6	± 7.8

We also looked at the formal errors of the baselines. For this evaluation the re-scaled solutions were used. Figure 4 shows the mean baseline error relative to the baseline length. The dotted line shows the baseline errors of the GIPSY/OASIS II solution while the dashed line shows the errors for the Bernese type solutions. It should be mentioned that the DGFI and IBGE solutions show the same behaviour and are therefore summarised as the Bernese type solution.

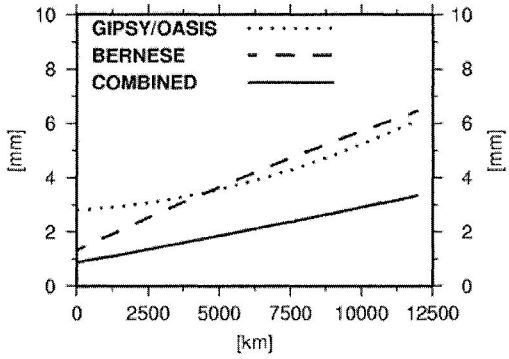


Fig. 4: Baseline length standard deviations relative to the baseline length.

Figure 4 shows quite clearly the difference between the precise point positioning approach and the relative positioning strategy of the Bernese software. Common errors in the data of neighbouring stations cancel out when applying relative positioning. The correlation of the data leads to higher precision between neighbouring stations. This is not the case for precise point positioning. Here the formal error of the baseline components increases with $\sqrt{2}$ of the coordinates components. Nevertheless, with increasing length (approx. 5000 km) the formal error of the baseline length approaches the formal error of the relative positioning. One should keep in mind that figure 4 is based on the re-scaled solutions.

4 Comparing 1995 and 2000 Results

To compare the results for the coordinates determination with those of 1995, the 2000 coordinates were transformed from ITRF2000 to ITRF94 where the 1995 results are referred to. The official ITRF transformation parameters were applied. The coordinates differences were divided by 5 years in order to get linear velocities and to make them comparable with the velocities derived from the IGS Regional Network Associate Analysis Centre for SIRGAS (DGFI02P01). The comparison is graphically illustrated by figure 5.

We see the quite homogeneous velocity field of the stable part of the South American plate which coincides well with the NNR NUVEL-1A model and the different behaviour of velocities in the Andean deformation zone. The average difference in 20 comparable stations included in both SIRGAS campaigns and the DGFI02P01 solution is +2 mm/a in latitude, −1 mm/a in longitude and 2 mm/a in height. After removing the systematic difference we get a deviation of ±1.2 mm/a in latitude, ±2.3 mm/a in longitude and ±4.3 mm/a in height. There are some stations with obviously large discrepancies in height (Bogotá, Maracaibo, Easter Island) where the 1995 and 2000 estimates differ by up to 5 cm.

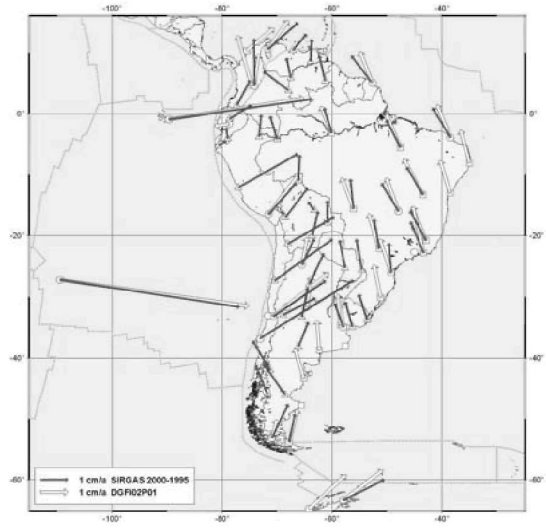


Fig. 5: Comparison of horizontal velocities from SIRGAS 1995 to 2000 and IGS RNAAC-SIR

5 References

Costa, S.M.A., E.S. Fonseca Junior, J.A. Fazan, J.F.G. Monico, P.O. Camargo (2002): Preliminary results of SIRGAS 2000 campaign. IBGE Analysis Center. IAG Symposia, Vol. 124, 306-311.

Drewes, H., L. Sánchez, D. Blitzkow, S. de Freitas (2002): Scientific foundations of the SIRGAS vertical reference system. IAG Symposia, Vol. 124, 297-301.

Hugentobler, U., S. Schaer, P. Fridez (Eds.) (2001): Bernese GPS software version 4.2. Astronomical Institute, University of Berne.

Kaniuth, K., H. Tremel, H. Drewes, K. Stuber, R. Maturana, H. Parra (2002): Processing of the SIRGAS 2000 network at DGFI. IAG Symposia, Vol. 124, 312-317.

Luz, R.T., L.P.S. Fortes, M. Hoyer, H. Drewes (2002): The vertical reference frame for the Americas – The SIRGAS 2000 GPS campaign. IAG Symposia, Vol. 124, 302-305.

Niell, A.E., (1996): Global mapping functions for the atmospheric delay at radio wavelengths. J. Geophys. Res. 101 (B2), 3227-3246.

SIRGAS (1997): Final Report Working Groups I and II. IBGE Brazil, Rio de Janeiro.

Webb, F.H., J.F. Zumberge (1997): An introduction to GIPSY/OASIS II. JPL Publication D-11088.

Zumberge, J., M. Heflin, D. Jefferson, M. Watkins, F. Webb (1997): Precise point positioning for the efficient and robust analysis of GPS data from large networks. J. Geophys. Res. 102, pp. 5005-5017.

Implementation of a Semi-Dynamic Datum for New Zealand

G. Blick, C. Crook, D. Grant

Land Information New Zealand, Lambton House, 160 Lambton Quay, Private Box 5501, Wellington, New Zealand; fax: +64 4 460 0112; email gblick@linz.govt.nz

J. Beavan

Institute of Geological and Nuclear Sciences, PO Box 30368, Lower Hutt, New Zealand

Abstract. New Zealand Geodetic Datum 2000 (NZGD2000) is a 'semi-dynamic' datum in that coordinates are published in terms of their values at the reference epoch of 1 January 2000 and it incorporates an as yet unpublished deformation model which allows positions at other times to be extrapolated. The proposed deformation model has two components, 1) a national deformation component using a latitude/longitude grid which is spatially and temporally continuous, and 2) a number of 'patches' used to model specific deformation events such as earthquakes. The national model will comprise a series deformation grids at fixed time intervals between which the deformation can be interpolated. The patches are localised triangulation based models of limited extent and time that can represent arbitrarily complex deformation. This is seen as meeting both the needs of the geodetic community, who require accurate coordinates at arbitrary times, and the mapping community, who prefer static coordinates, but expect them to reflect major deformation events.

Keywords. Datum, velocity model, deformation, New Zealand

1 Introduction

In 1998 Land Information New Zealand (LINZ) implemented a new geocentric datum for New Zealand, New Zealand Geodetic Datum 2000 (NZGD2000) with a reference epoch of 1 January 2000 (2000.0). A major conceptual departure from the definition of the previous national datum (New Zealand Geodetic Datum 1949) and other international datums is that NZGD2000 accommodates the effects of crustal deformation. This is achieved by applying a deformation model when generating new coordinates. This enables coordinates to be transformed from one epoch to another following a similar method described by Snay (1999). For many users, it has the appearance of a static datum, which facilitates use of the datum by the GIS and mapping community who do not have the tools to manage constantly changing coordinates. (In the future, a fully dynamic datum will provide for the deformation model to be made generally available to users to generate coordinates at any user-specified epoch).

NZGD2000 is realised in terms of ITRF96 and uses the GRS80 ellipsoid, see Grant and Blick (1998) and Grant et al (1999). Coordinates for 29 primary 1st Order 2000 network stations were generated from 5 repeat GPS surveys made between 1992 and 1998 (Office of the Surveyor-General, 2000). Data from these and other repeat surveys were used to generate a horizontal deformation model by the Institute of Geological and Nuclear Sciences (GNS) (Beavan and Haines, 2001). The deformation model (Figure 1) is used to account for broad scale crustal deformation across New Zealand, primarily due to the effects of plate tectonics.

The current deformation model used for the definition of NZGD2000 assumes a constant deformation velocity through time. The surveys used to determine the deformation model are now five years old. As time passes, errors in the determination of the velocities used in the deformation model will lead to increasing errors in the calculated position of marks in terms of the reference epoch of 2000.0. In effect, the spatial accuracy of the datum is steadily degrading. Additionally, the datum and effectiveness of the deformation model may be degraded by localised and temporally non-linear deformation, eg earthquakes. The following issues are now considered:

- How much error can be tolerated before a new deformation model is required?

- When there is local and spatially complex deformation, how much should be accommodated by the deformation model, and how much should be accommodated by changing the reference epoch (2000.0) coordinates?
- What temporal model should be used to accommodate non-linear changes in deformation?

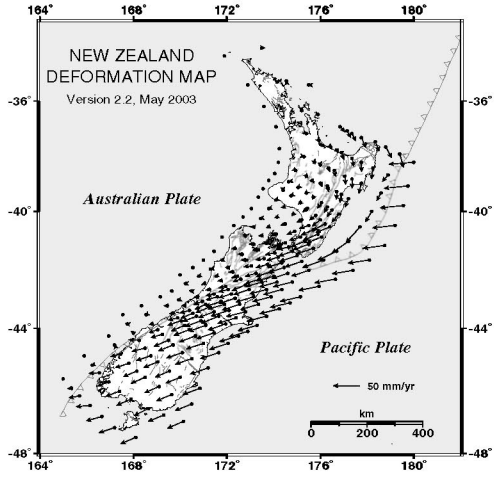


Fig 1. New Zealand Deformation Model (Beavan and Haines, 2001).

This paper considers these issues and presents New Zealand's option for implementing a semi-dynamic datum. A full description of the various options that were considered is available in Office of the Surveyor-General (2003a).

2 How Much Error can be Tolerated in the Deformation Model?

A general principle in the development of NZGD2000 is that: '*the accuracy of a mark's coordinates relative to adjacent marks of the next highest order shall be dependent on the distance between them and shall not exceed 0.05m horizontally and 0.15m vertically*' (Office of the Surveyor-General 2003b).

It follows that the coordinate error and deformation model must be of sufficient accuracy to enable these accuracy requirements to continue to be achieved over time. Where the computed positions of marks using the deformation model differ from the surveyed positions by greater than these limits, consideration will need to be given to refining the deformation model.

3 Spatial Definition of Deformation Model

The spatial definition of deformation must be able to reflect the true deformation field with an adequate resolution. A model should include both the long term deformation trends and potentially discrete events such as earthquakes, where the model definition could include surface fault ruptures. This would depend upon the extent to which fault movement should be reflected by the deformation field, and the extent to which it should be represented by changing the coordinates of survey marks.

To reflect this spatial complexity the deformation model must be defined by dividing the area covered into a number of regions, and defining interpolation functions within each region. Three options for this breakdown were considered (Office of the Surveyor-General, 2003a):

- Simple rectangular grid
- Complex grid (eg curvilinear grid)
- Triangulated or other irregular grid

The simple rectangular grid was selected for the national deformation model, as it is well understood and there are standard formats and implementations of grids. It is also very efficient for calculating deformation at any point on the grid (mainly because it is very simple to identify which grid cell a point lies in). However, where discrete events such as earthquakes occur, more complex models may be required, and a triangulated grid may be more appropriate (see section 5).

4 Temporal Definition of Deformation Model

The current deformation model defines a constant velocity at each point. This will not remain valid for long, however, as:

- The model is based on data of limited spatial and temporal extent. Over time the imperfections in the model will become apparent and will require a revised model (though potentially still based a constant velocity);
- The deformation field itself may not be constant; and
- There may be discrete deformation events (eg earthquakes) which are not representative of, or represented by, the long-term trends.

It therefore follows that the model will require periodic updates. The following models were considered (Office of the Surveyor-General, 2003a):

Steer to the new model: When new information is available the mark is 'steered' back towards its believed position and velocity over a period of time.
Jump to the new model: The original model is preserved. The new deformation model replaces the original for all future times. A discontinuity is introduced at the time of model update.
Revise the previous model: The new information is used to revise the previous model. The new model may not necessarily have constant velocity.
Ignore previous model: The new information is used to calculate a new constant velocity deformation model. The old model is discarded. Reference epoch coordinates are updated to reflect the backward extrapolation of the new model.

These models differ in the extent to which coordinates and velocities prior to the update are allowed to change. A key consideration for a national geodetic datum is the need to maintain the coherence of coordinates for a range of users over an extended period of time. Therefore, both the spatial and temporal continuity of the model are important.

Of the options considered, the third option, **Revise the previous model**, is adopted (Figure 2).

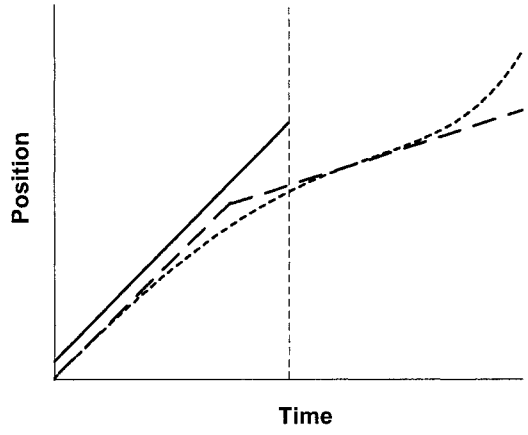


Fig 2. One dimensional position of a mark against time. Dotted line shows the actual movement of the mark. The solid line is the initial deformation model and the dashed line is the revised deformation model after the time of updating, indicated by the vertical dashed line.

In such a model the original model is not necessarily retained, though older deformations are unlikely to change, and may be deliberately preserved. In this option the most recently published model always represents the best estimate of past and future deformation.

5 Accommodation of Discrete Events

Earthquakes, major landslides, and other discrete deformation events are ignored by the current deformation model. Where the event involves spatial discontinuity, such as a fault rupture, it may be difficult to handle with a deformation model at all. Where and when the deformation model fails to match the actual deformation, the survey marks are effectively moved to new logical locations – in effect the old mark appears in a new position. This complicates management of the survey network.

Three potential options for representing deformation events in the datum were considered (Office of the Surveyor-General, 2003a):
Densify the model: Redefining the national deformation model with a higher density of points, at least in the vicinity of the deformation event. This is only sensibly possible with the triangulated model, which is not the preferred option for the national model.
Define a local 'patch' for the model: Publish a local perturbation to the national deformation model. To ensure spatial continuity of the total model (ie national model plus patches) the perturbing model would have zero deformation at its boundaries (except where it extends to the edge of the national model). To calculate the deformation at a given time and place would require identifying which patches, if any, apply and adding the deformation from them to that from the national model.
Change coordinates: With this option the national model would remain unaltered, but the coordinates of all points influenced by the local event would be updated. This would ensure that deformation calculations remained simple, but may require some form of versioning of marks and coordinates. Where property boundary marks are managed in terms of the geodetic system (as in New Zealand), a very large number of marks may be affected.

Of these options the use of a patch is considered most practical. A patch is essentially a localised deformation model defined over the area and time for which the deformation perturbation exists. For an earthquake, the area of the patch would be the region around the epicentre in which significant coseismic and post-seismic deformation was detectable. The model may define ongoing post-seismic deformation for a period after the earthquake during which, anomalous post-seismic deformation is detectable. The model would also define a permanent 'final' offset resulting from the earthquake, which would apply to all coordinates after the event. In order to model the complexity of deformation associated with local deformation events, the patch

would need to use a triangulated or similar irregular model.

If we use patches then the existing deformation model will be able to reflect discrete events, at least as well as we are able to measure them. The reference epoch coordinates could be retained, and we will be able to combine observations before and after the event in a single adjustment in which the differences due to deformation are fully accounted for by the deformation model. A difficulty with this approach is that the official coordinates of the datum, in this case the 2000.0 coordinates, will not closely reflect the current relative positions of the marks. For example, adjacent marks on opposite sides of a fault line may have moved metres relative to one another during an earthquake. The 2000.0 coordinates without the patch model applied will give relative positions that do not fit current observations. Users will only find coordinates useful if they have applied the patch to derive them. Use of the deformation model and its patches thus becomes virtually essential for all users.

What is required is datum and deformation models that meet two requirements – firstly they model the discrete event well enough that old and new observations can be used together, and secondly that they easily provide users with coordinates that are sensible in terms of the current positions of marks. The proposed approach to handling this is to use patches 'in reverse'. Instead of using patches to correct unchanging base or reference epoch coordinates to their current positions, we will update the reference epoch coordinates to reflect the effects of the perturbing deformation, and use the patch as a 'negative' deformation to determine coordinates for times before the deformation event.

The principal advantage of this approach is that calculating current coordinates (the coordinates most often required by users) does not involve calculating the patch deformation – it is simply defined in terms of the base epoch coordinate and the (relatively simple) national deformation model. However most coordinates remain static, which is a user requirement of most GIS and mapping users. The only coordinates that change are those directly affected by earthquakes and similar events, and most users would intuitively expect these to change.

The implementation of the 'patch' is illustrated in the following two diagrams. In Figure 3 the deformation event is not incorporated into the deformation model. The dotted line is the actual (one dimensional) position of the mark plotted against time and the dot the base epoch coordinate for the mark. The mark is affected by a discrete deformation event. The line is the trajectory of the mark defined by the base epoch coordinate and the national deformation model. The datum (base coordinate plus deformation model) does not represent the position of the mark after the deformation event.

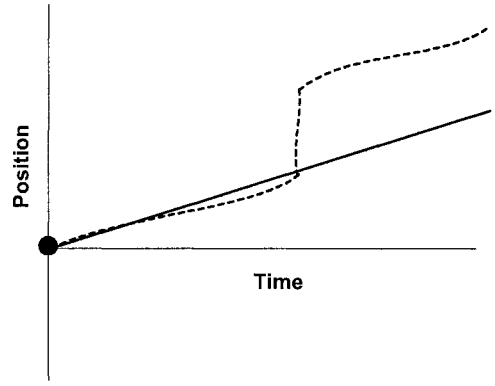


Fig 3. Coordinate of mark defined by deformation model that does not include the deformation event.

In Figure 4 the deformation event is incorporated into the deformation model. The long dashed line is the patch deformation model, which in this case is simply a fixed offset at about the time of the event. The short dashed line shows the trajectory of the mark as defined by the deformation model, which includes the patch. The base epoch coordinate is changed to include the offset calculated from the patch. The patch deformation model is non-zero at the base epoch, and so the datum trajectory (base epoch coordinate plus total deformation model) is the same as the model above. The patch deformation is zero after the event, so that coordinates for times after the event may be simply calculated from the new base epoch coordinate and the national deformation model.

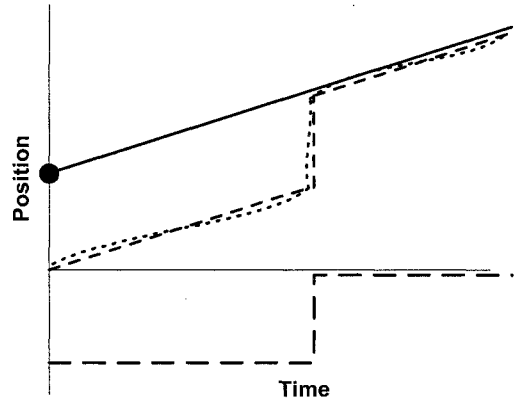


Fig 4. Coordinate of mark defined by deformation model incorporating a patch to model the deformation event.

Not all discrete events will be able to be represented by an instantaneous offset at a specific time. Earthquakes may be followed by significant post-seismic deformation continuing for months or even years after the event. Where there is ongoing anomalous deformation it may still be adequate to represent the event with a discrete offset at a specific time. If there are no surveys significantly affected by the ongoing deformation (eg, because all such surveys are either before or after the period of deformation) then there is no need to model that deformation. The most that may be required is to revise the patch model and the affected base coordinates if the initially published models do not reflect the total deformation.

If it is necessary to model the ongoing deformation then the patch model will need to include more than one epoch. When each new version of the patch is defined it is likely to redefine the total deformation from the event, which means that there will need to be a corresponding adjustment to the base coordinates.

6 Other Issues to Consider

6.1 Accommodating Vertical Deformation

The current model only defines horizontal deformation. However vertical motion, although generally smaller than horizontal motion, is known to exist and this motion must be accommodated in the future. Unfortunately the national vertical deformation trends are in many places obscured by much larger localised episodic or cyclic events, such as response to geothermal steam extraction, drainage, and so on. This means that in many areas the datum will not be able to reflect the vertical deformation well, and even using a patch type of approach may be too complicated to realise. Nonetheless the national trends should be discernable by careful selection of the survey marks used to define it. This will be integrated into the national model by simply including vertical as well as horizontal deformation components.

6.2 Latency

There may be considerable delay between a deformation event (or a large scale variation in the deformation field) and it being detected, measured, and implemented in a new deformation model. As New Zealand's national network of continuously operating GPS receivers densifies, this latency may decrease, at least for the national model. For discrete events the deformation will often be complex, and our initial understanding of it may be incomplete. Patch models may require several versions as new information becomes available.

6.3 Confidence

The deformation model will have regions and times of differing confidence. Also our confidence will reduce when extrapolating well beyond the period over which the data generating the model was observed. This uncertainty and other appropriate metadata should be published with the model.

6.4 Extension Offshore

The current velocity model only provides deformations for the New Zealand land area. If we wish to calculate deformations for offshore locations we will need to include a definition of a more extensive model (possibly a global model). For a global model it may be preferable to express the velocity as a global rotational velocity, rather than east and north components of velocity.

6.5 Changing the Reference Epoch

Ultimately there will come a time when coordinates referenced to epoch 2000.0 become inconveniently different from the true current positions of coordinates to the extent that the difficulties of managing the differences outweigh the desire of many users for static coordinates. At that time the national deformation model can be used to migrate the coordinates to a new reference epoch – effectively a new datum.

7 Proposed Deformation Model

The proposed implementation of a deformation model for NZGD2000 is as follows:

1) The deformation model will comprise a national ongoing deformation model and an arbitrary number of 'patches' representing specific deformation events.
2) The national component will define the deformation at specific epochs (eg 2000, 2005, 2010). The time between epochs is to be determined. Deformation at other times is determined by linear interpolation between epochs.
3) The national component will include two deformation models, one for extrapolating times before the first epoch, and one for extrapolating times after the last epoch (including the present).
4) The deformation models will be defined on a rectangular grid (in terms of latitude and longi-

tude). The deformation will be interpolated within the grid by bilinear interpolation methods.

5) When a new deformation model is published, it may redefine any of the previous epoch deformation models if there is new information to justify doing so.
6) Specific deformation events, such as earthquakes, will be added to the model as 'patches', which represent the perturbation to the deformation field due to the event. The patch models will be defined to span the spatial extents of the significant and measurable deformation. The deformation on the boundary of the patch model will be zero (except where it is also the boundary of the national model).
7) The patch model will define the deformation before the event relative to the current position – in effect the patch is a negative deformation event. The base epoch coordinates of all affected marks will be updated to reflect the deformation due to the event. The patch models will be assumed to have zero velocities before and after the event.
8) Where an event includes ongoing anomalous deformation the patch model may include several epochs between which the deformation will be interpolated. The models will always have a final deformation of zero. This may mean that several versions of the patch are published as new information is obtained.
9) When a new patch model is published there will be a corresponding update to the coordinates of affected marks to represent the revised total effect of the deformation on the coordinates. New models may be published either because there has been a new deformation event, because there has been ongoing deformation of a previous event, or because better information has become available about a previous event.
10) Each patch model will be based on a triangulated network in order to be able to represent arbitrarily complex deformation such as localised deformation across a fault trace.
11) Both the national component and the patches may define horizontal and vertical deformation.

8 Summary

The method described here differs from that proposed by Snay (1999) in that it is a purely kinematic model that does not attempt to model earthquake dynamics. Though scientifically less appealing, this makes it more flexible to implement, update and use. Though the definition here is not expressed in terms of dislocation models, such models may be used to calculate the deformation model.

The current deformation model using a constant velocity and implemented in NZGD2000 provides a short term solution to manage the effects of crustal motion on the datum. However as we move further away from the datum's reference epoch, 2000.0, a number of limitations with this simple model will become apparent. A deformation model has been proposed that will overcome these limitations and ensure that NZGD2000 remains a modern accurate geodetic datum that serves New Zealand's spatial needs into the future.

9 References

Beavan, J, and J. Haines, Contemporary horizontal velocity and strain-rate fields of the Pacific-Australian plate boundary zone through New Zealand, *J. Geophys. Res., 106:B1, 741-770,* 2001.

Grant, D.B., G.H. Blick, M.B. Pearse, R.J. Beavan, and P.J. Morgan, The development and implementation of New Zealand Geodetic Datum 2000. *Presented at IUGG 99 General Assembly, Birmingham UK, July 1999*, 1999.

Grant, D.B. and G.H. Blick: A new geocentric datum for New Zealand (NZGD2000), *New Zealand Surveyor, No. 288. pp 40-42,* 1998

Office of the Surveyor-General, Implementation of the NZGD2000 Velocity Model, *Land Information New Zealand, Office of the Surveyor-General Technical Report 20.* Available from www.linz.govt.nz/surveypublications, 2003a

Office of the Surveyor-General, Accuracy Standards for Geodetic Surveys, *Land Information New Zealand, Office of the Surveyor-General Standard 1, v1.1.* Available from www.linz.govt.nz/surveypublications, 2003b

Office of the Surveyor-General, Realisation of the New Zealand Geodetic Datum 2000, *Land Information New Zealand, Office of the Surveyor-General Technical Report 5.* Available from www.linz.govt.nz/surveypublications, 2000

Snay, R.A.: Using the HTDP software to transform spatial coordinates across time and between reference frames, *Surveying and Land Information Systems, Vol. 59, No. 1, pp. 15-25, 1999*

GPS measurements to investigate the reason why GPS is less accurate in mountain areas

M. Satomura[1], S. Shimada[2], Y. Goto[1,3] and M. Nishikori[1,4]

1 Faculty of Science, Shizuoka University, Oya, Shizuoka 422-8529, Japan
e-mail : semsato@ipc.shizuokja.ac.jp (M.S); Tel: +81-54-238-4945; Fax: +81-54-248-9031
2 National Research Institute for Earth and Disaster Prevention, Tennodai, Tsukuba, Ibaraki 305-0006 Japan
3 Present affiliation : Kanazawa Branch, Japan Travel Bureau
4 Present affiliation : Nippon GPS Solutions Corporation

Abstract. It has been pointed out that GPS results in mountain areas are usually less accurate than those in flat areas, especially in the vertical component, when the height difference among observation stations is large. It is thought that the decrease of accuracy is caused by insufficient corrections for tropospheric delay of the GPS microwave.

In order to investigate the difference of tropospheric delays at sites located on high mountains and those at valleys, and furthermore to improve the accuracy of GPS results in mountain areas, we have repeated GPS measurements since September 1997 at Yamainudan (about 1400 m above sea level) near the top of Mt. Sobatsubu in the Southern Alps of Japan. Their data were processed with the data at four stations of surrounding valleys (300 - 400 m) by using GAMIT software, to investigate the relation of the coordinates obtained and zenith tropospheric delays. We found a negative correlation between the difference of zenith tropospheric delay and height difference between Yamainudan and a site at valley. These results scarcely changed when a mapping function used was changed.

We also processed some GPS data obtained by the Geographical Survey Institute (GEONET data) in the mountain areas. A positive correlation between the difference of tropospheric zenith delay and height difference was found in the Northern Alps, and no significant correlation was found in the Central Alps of Japan.

We think that these different correlations were caused by the difference of the shape of the valleys.

Key word; GPS - Vertical component - Mountain area

1. Introduction

The accuracy of the GPS measurements has achieved very accurate positioning in horizontal components by correcting the horizontal gradients of the atmospheric delay, but it is not very accurate in the vertical component yet. And furthermore, it has been pointed out that GPS observation in mountainous areas is usually less accurate than that in flat areas, especially in the vertical component when the height difference among observation sites is large. It is thought that the decrease of the accuracy is caused by insufficient corrections for tropospheric delay of the GPS microwave.

In order to investigate the difference of tropospheric delays at sites located on high mountains and those at valleys, and furthermore to improve the accuracy of GPS observation in mountain areas, we repeated GPS measurements at Yamainudan near the top of Mt. Sobatsubu in the Southern Alps of Japan. Its height is about 1400 m above sea level. Their data were processed with the data at five stations located along valleys surrounding Mt. Sobatsubu and also with semi-global IGS data, to investigate the relation of the coordinates obtained and zenith tropospheric delays (ZTD).

We also processed some GPS data obtained by the Geographical Survey Institute (GEONET data) in the mountain areas. A positive correlation between the difference of zenith tropospheric delay and height difference was found in the Northern Alps to compare with the results obtained in the Southern Alps.

2. Observation in the Southern Alps of Japan

Permanent GPS stations called GEONET have been constructed by the Geographical Survey Institute all over Japanese islands, but their stations are usually

located along valleys in mountain areas, because of convenience of access to the stations. In the Japanese Southern Alps, central part of Japan, GPS stations are also located along valleys, for example, Minami-Shinano, Misakubo, Haruno-North, Honkawane-South, and so on. Their elevations are about 300 m or 400 m above sea level, although there are many high mountains whose heights are sometimes more than 3000 m. Mt. Sobatubu is located in the southernmost part of this mountain area, and we can access to the altitudes up to 1400 m by car. We made a GPS observation site there, and performed some observation campaigns since September 1997. Sampling interval is 30 sec. A map showing the observation sites used in the present investigation is shown in Fig. 1.

We processed their data with four stations data of GEONET by the Geographical Survey Institute, by using the GAMIT software. All data processed in the present investigation were obtained by Trimble 4000 SSE or 4000 SSi receivers.

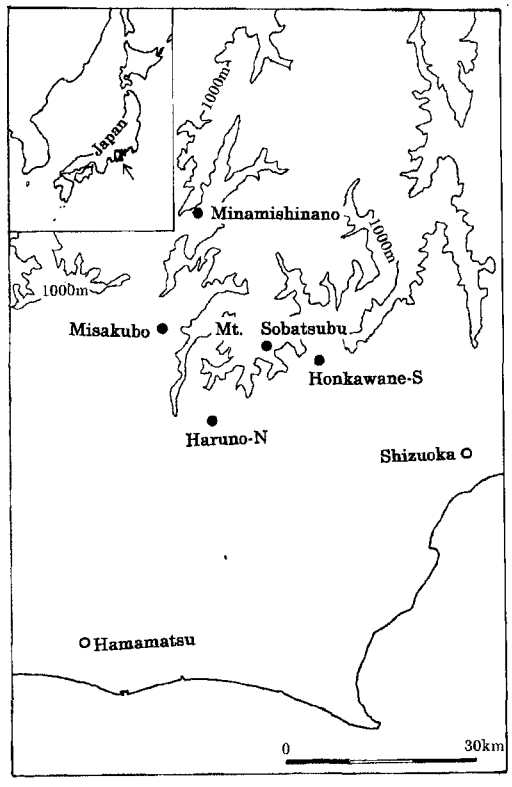


Fig. 1. GPS stations whose data are used in the present investigation in the Southern Alps of Japan. We performed GPS observations at Mt Sobatsubu as campaign observations, and the Geographical Survey Institute preformed the continuous observations at the other stations as GEONET.

3. Data Processing and Results in the Southern Alps of Japan

The data obtained were processed to obtain two kinds of results; one is geodetic coordinates of the sites and the other is zenith tropospheric delays (ZTD).

First, we determined the coordinates of their sites by processing referring to five IGS stations; Fairbanks, Shanghai, Kokee Park, Tsukuba and Usuda by using GAMIT and GLOBK (Herring et al., 1990) software. By constraining the coordinates to be less than 5 mm for three components, we processed again the data to obtain the ZTD for each one hour.

Some of these results were checked with the water vapor change data obtained by water vapor radiometer and radiosonde

When GPS measurements for crustal movement studies are performed, coordinate differences among some GPS sites are usually obtained by processing the local GPS data referring to a certain base station. Therefore, we also calculated coordinates of the stations only from local station data by fixing the coordinate of Honkawane-South, and the calculated ellipsoidal height of Mt. Sobatsubu.

We compared the ellipsoidal height difference with the difference of zenith tropospheric delay. Fig. 2 is an example of the comparison and it is that between Mt. Sobatsubu and Honkawane-South when the coordinates are obtained by referring to IGS stations. Fig. 2 shows that there is a negative correlation between the difference of zenith tropospheric delay and the height difference.

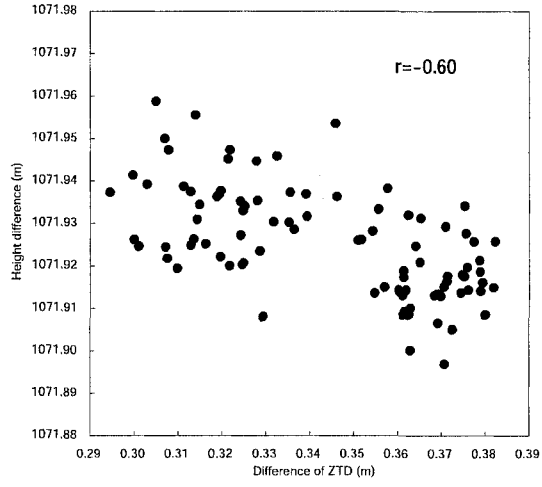


Fig. 2. Relation between the height difference and ZTD difference between Mt. Sobatsubu and Honkawane-south in the case that the coordinate was obtained by referring to IGS stations. The correlation coefficient is -0.60.

This result shows that we may make over-correction about tropospheric effects in the data procession.

These results were obtained by processing with NMF mapping function. (Niell, 1996). We also processed the data by using other mapping functions, i.e. CFA, MARI and CHAO mapping functions. The relation between the difference of the zenith delay and height difference slightly changed and this negative correlation still remains.

4. Results in the other areas

In order to investigate whether the results which we obtained from the data in the Southern Alps of Japan were general or not, we processed some GEONET data obtained in the Central and Northern Alps of Japan by the same processing as in the Southern Alps. The location of the stations used in this processing is shown in Fig. 3 as well as that in the Southern Alps.

The obtained correlation coefficients between height difference and ZTD difference are shown in Table 1. The results in the Northern and Central Alps are much different from those in the Southern Alps. There are positive relations in the baseline of the Northern Alps, and there are no significant relations in the Central Alps.

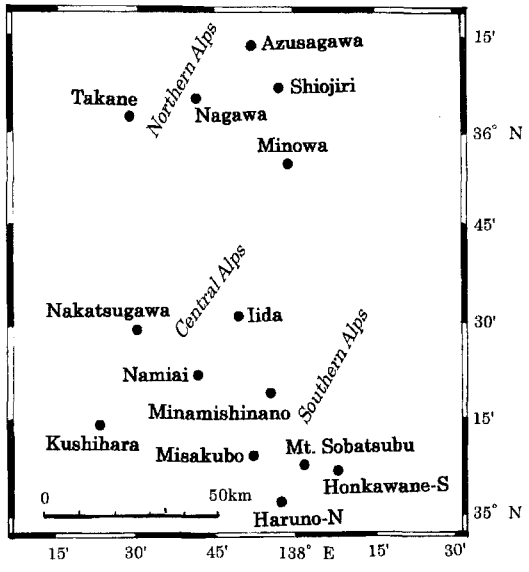


Fig.3. Location of the all GPS sites processed in the present investigation. Azusagawa, Takane, Nagawa, Shiojiri and Minowa are in the Northern Alps; Iida, Nakatsugawa, Namiai and Nakatsugawa and Minamishinano are in the Central Alps; and Minamishinano, Misakubo, Mt. Sobatsubu Honkawane-S and Haruno-N are in the Southern Alps.

Table 1. The obtained correlation coefficients between height difference and ZTD difference

Southern Alps	Semi-Global	Local
Minamishinano - Mt. Sobatsubu	-0.17	-0.29
Haruno-N - Mt. Sobatsubu	-0.59	-0.54
Misakubo - Mt. Sobatsubu	-0.51	-0.44
Honkawane - Mt. Saotsubu	-0.60	-0.54
Central Alps		
Nakatsugawa - Namiai	-0.22	0.08
Iida - Namiai	-0.17	-0.21
Minamishinano - Namiai	-0.23	-0.17
Kushihara - Namiai	-0.25	-0.16
Northern Alps		
Minowa - Takane	0.18	0.04
Shiojiri - Takane	0.64	0.17.
Azusagawa - Takane	0.44	0.13
Nagawa - Takane	0.49	0.42

5. Discussions

In order to investigate the reason why such different results are obtained in the three areas, we studied three points which may cause the difference.

1 Effect of the height difference

We studied the relationship between the obtained coefficients and height differences. When the height differences are less than 200 m, there are no relations between the obtained height difference and ZTD difference. There are positive relations when the height difference is 200 - 800m, and there are negative relations when the height difference in more than 1000 m. Only the baselines in the Southern Alps have more than 1000 m height differences, and therefore we cannot decide that this is caused by the difference of height difference or area difference.

2 Load effect by snow

Northern Alps has much snow in the winter season, but Southern Alps has only a little snow. Heki (2001) showed that snow load causes the baseline length change up to 1 - 2 cm in the northeastern Japan, and therefore the load may affect the present results.

We check the Azusagawa - Takane baseline in the Northern Alps which has a positive relation in detail. The height differences obtained in November when there is little snow coverage are almost the same as those in January and March when there is much snow coverage.

This fact shows that the relation obtained in the present research is not made by snow load.

3 Width of valley

We investigate the relation between the obtained coefficients and width of the valley where the lower site of the baseline is located. The width of the valley is read from map where the height difference is 500 m from the lowest point of the valley. The result is shown in Fig. 4. There is a positive relation in Fig. 4. We also make another figure which we read the width where the height difference in 300 m and obtained almost the same result.

The water vapor change must be different in the narrow valley and wide valley, but we process the GPS data in the same condition. This fact may show that we must change the method to assume the excess path delay of GPS microwave by the shape of the valley when we process GPS data in mountainous area.

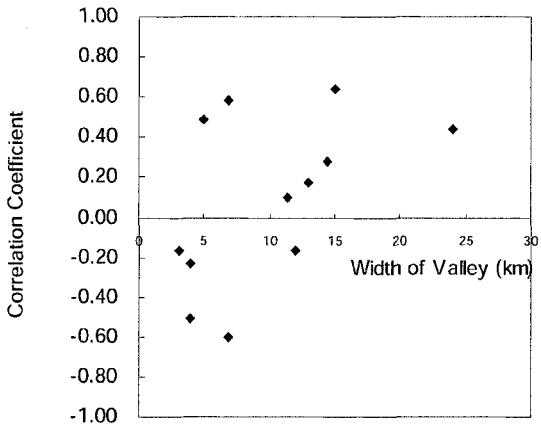


Fig. 4. Relation between the width of valley where the lower site is located and correlation coefficients between height difference and ZTD difference.

5. Conclusions

In order to investigate how we can improve the accuracy of GPS observation in mountain areas, we process the GPS data in the mountainous areas of central Japan with GAMIT software.

A negative correlation between the height difference and ZTD difference in the baseline was obtained in the Southern Alps, but the results in the Central and Northern Alps are much different from the Southern Alps one.

More detailed examinations are needed to make clear the cause of this difference, but we think in the present stage that it may be caused by the difference of valley shape. If so, we must think a suitable processing method associated with the shape of valley.

Acknowledgements

The authors wish to express their sincere thanks to the Geographical Survey Institute to provide them the GEONET data. They also thank the Meteorological Research Institute for observing the water vapor change concerned with a radiometer and radiosondes.

References

Niell, A. E., Global mapping function for the atmospheric delay at radio wavelengths. J. Geophys. Res., 101: 3227-3246. 1996.

Herring, T. A., Davis, J. L., Shapiro, I. I., Geodesy by radio interferometry: The application of Kalman filtering to the analysis of Very Long Baseline Interferometry data, J. Geophys. Res., 95: 12561-12581, 1990.

Heki, K., Seasonal modulation of interseismic strain buildup in northeastern Japan driven by snow loads. Science, 293: 89-92, 2001

Towards More Reliable Estimation of GPS Positioning Accuracy

J. Krynski, Y.M. Zanimonskiy
Institute of Geodesy and Cartography, 27 Modzelewskiego St., 02-679 Warsaw, Poland
e-mail: krynski@igik.edu.pl, yzan@poczta.onet.pl; Fax: +48 22 3291950, Tel.: +48 22 3291904

Abstract Uncertainty of vector components estimation, obtained from processing GPS data using either commercial or scientific software, represents rather an internal consistency than the accuracy of positioning. Short period biases, including non-modelled effects, some with unstable amplitudes, that affect GPS solutions are as systematic-type terms not reflected in uncertainty estimation. Therefore, calculated uncertainty is usually much too optimistic as the accuracy estimate. In the routine GPS data processing one observing session provides a single solution. Data that form such observing session can be represented by a series of overlapped sessions. Comparison of solutions obtained from processing overlapped sessions leads to more reliable accuracy estimation than the one based on uncertainty estimation of individual solutions. The strategy of GPS solutions quality analysis based on the concept of overlapping sessions with optimum length and temporal resolution is presented. The strategy was verified with use of data from the EPN stations processed with both Bernese and Pinnacle software packages.

Keywords: Global Positioning System (GPS), Positioning accuracy, Statistical analysis, Spectral analysis

1. Introduction

It is well known that the solutions for vector components or coordinates obtained from processing precise GPS data from different observing sessions vary usually much stronger than what their precision estimate indicates. Standard deviations of GPS solutions provided by processing software reflect only the internal consistency of data processed and the internal precision. They do not, however, indicate the actual accuracy of GPS positioning in the real scale (e.g. Gandolfi et al., 2003).

An extensive research is conducted to improve the precision of GPS solutions by using better models for GPS observations, and to improve the precision estimate (e.g. Teunissen, 2002). Time series of GPS solutions are particularly suitable for such investigations (Krynski and Zanimonskiy, 2002). The tools of statistical analysis and spectral analysis are useful for separation factors causing variations in GPS solutions and estimation of their magnitudes.

Time series of GPS solutions do not exactly represent a random process. Besides a random part that is mainly due to observation noise the variations in GPS solutions also contain components of a chaotic character as well as biases (Krynski et al., 2002a). Model errors, non-modelled effects and varying satellite configuration, including multipath cause systematic variations in GPS solutions. In addition, due to non-linearity of the system, data noise generates biases in computed results. Missed cycles in integer ambiguity resolution process and sudden changes in satellite configuration due to a rise of a satellite or satellite's repair are the main sources of chaotic errors.

The choice of the method used to suppress disturbances in time series corresponds to their character, i.e. random or chaotic. Noise, external with respect to measuring system, is filtered out using the smoothing procedure that employs a rectangular window the size of which becomes a parameter to be determined. The optimum size of the window can be estimated at each filtering stage. The analysis of such time series indicates the existence of a number of periodic components and trends not modelled in data processing stage (Bruyninx, 2001; Poutanen et al., 2001; Krynski et al., 2002a). Major part of the power spectrum of the variations is concentrated in diurnal and longer periods. In particular, periodic variations with dominating 12h and 24h periods (Krynski et al., 2000) are distinguished. Few hours' long periods in the spectrum are most probably the artefacts (King et al., 2003) caused by the effects dominated by random noise due to non-linearity of the system (Krynski and Zanimonskiy, 2002).

Processing of as long as 24h GPS sessions leads to solutions smoothed off for daily and sub-daily periodic biases. The use of shorter observing sessions with preserving high quality of GPS positioning requires the investigation of periodic biases, their detection, their source specification and an attempt towards their modelling.

Due to a large amount of information contained in time series of GPS solutions based on overlapped sessions it becomes possible to apply statistical tests to detect outliers. Numerous sudden changes in satellite configuration occur with a period of one sidereal day. They cause specific variations (jerks) in time series of GPS solutions, with periods significantly smaller than one sidereal day, i.e. even of the order of 1h. Such variations correspond rather to a chaotic process then a random one. Jerks can be suppressed by optimising the length of overlapping sessions and eliminating the disturbing results (Krynski and Zanimonskiy, 2002).

Continuous and regular change of satellite constellation causes smooth variations of parameters of measuring process, i.e. signal to noise ratio, atmospheric delays, multipath, orbit corrections, etc., that are periodic with a half of sidereal day period (Krynski et al., 2002b). Terms with periods of half and one sidereal day occur in informative parameters such as vector components as well as in non-informative parameters, like software-provided standard deviations of GPS solutions, cross-correlation coefficients of vector components, number of measurements accepted for the solution, etc. In time series, including those of GPS solutions, there exists a corresponding to the sampling interval periodicity that should be interpreted as the artefact. The shorter sampling interval the finer is time resolution, the higher is the Nyquist frequency and the less distinguished become spectral leakage effects in time series.

A careful estimation of an optimum length of a session used to calculate positions from GPS data is needed due to jerk-type variations in GPS solutions corresponding to a satellite rising or its descending. The optimum length of a session does not necessarily correspond to longer ones. With the increase of the session length an internal accuracy of the output data increases but at the same time the growth of spectral leakage is observed. Thus the extension of a session length used for computing GPS data reduces the estimated uncertainty of the solution but it simultaneously decreases Nyquist frequency. The sum of those two counteracting effects depends also on spectrum of noise and the signal itself. The increase of the Nyquist frequency can be accomplished by using overlapping sessions. Correlation accompanying time series of GPS solutions obtained from overlapping sessions is significantly smaller than the one in the classical random process series. For example, the correlation coefficient in time series based on solutions from the sessions with 87% overlap is at the level of 0.5 (Krynski et al., 2002b) while such a coefficient for a wideband random process reaches 0.5 in case of 50% overlap (Harris, 1978).

Predictability of the reaction of the GPS measuring system (both receiver and processing software) on disturbances and inadequacy of models used could be viewed experimentally by statistical and correlation analysis of time series of GPS solutions.

2. Numerical experiments

The problem of reliable accuracy estimation of GPS positioning can be investigated using time series of GPS solutions obtained from sessions of different lengths for vectors of different length, located in different geographic regions. Practical needs, potentiality of accessible data processing infrastructure, and the experience gained in GPS research contributed towards making the choice of GPS data for processing and determining its strategy for numerical experiments.

As an example, GPS data from 2001 provided by two EPN Stations BOGO and JOZE was used to generate time series of BOGO-JOZE vector (42 km length) components with the Bernese v.4.2 and Pinnacle software packages. With the Bernese software the GPS solutions were obtained from processing 2h, 3h, 4h, and 6h sessions over 19 days in August, with an overlap (1h shift), 3h sessions with 2.5h overlap (30m shift), and 24h sessions with 23h overlap over 4 months (February-May). GPS solutions were obtained with Pinnacle from processing 2h, 3h, 4h, 5h, 6h, 8h, 9h, 10h, 11h, 12h, 13h, 14h, 17h, 18h, 19h, 20h, 21h, 22h, 23h, 25h, 26h, 27h, and 28h with 1h time resolution (1h shift), over 15 days in August. Chosen data represent two different periods of seasonal atmospheric dynamics, i.e. winter-spring corresponds to quiet atmosphere while summer to a disturbed one. Time series of GPS solutions generated were then the subjects of statistical analysis. The dispersion of GPS solutions as well as their averaged combinations (mean values m, standard deviations σ_z and the estimated standard deviations σ_m of the mean) together with their precision estimates (software-provided standard deviations σ_w) was analysed. The conceptual scheme of forming groups of GPS solutions for further statistical analysis is shown in Fig. 1.

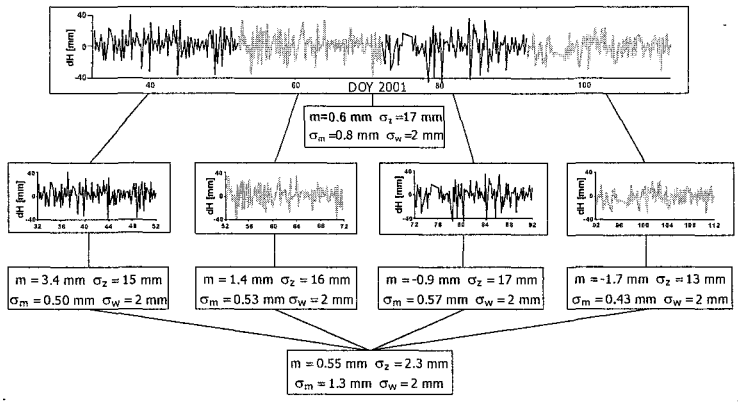


Fig. 1. The scheme of forming groups of GPS solutions and example of their estimated statistics

Variations of GPS solutions in the investigated time series are typical for the solutions based on data from numerous EPN stations. Standard deviations of those solutions that are the measure of internal accuracy are all at the same level. Thus mean standard deviation σ_w in the group of solutions in Fig. 1 and their combinations are all equal to 2 mm. Estimators of m, σ_z and σ_m that are used to estimate external accuracy of GPS solutions take different values in different groups. Averaging solutions that represent the groups (statistics in lowest rectangle in Fig. 1) provides similar result to the one based on all data included in the groups considered (statistics in upper rectangle in Fig. 1) but its accuracy estimate becomes more realistic (estimate of internal accuracy σ_w will correspond to the estimate of external accuracy σ_z).

Dispersion of GPS solutions grows with the size of sample, i.e. with a number of solutions that form the group investigated; that also corresponds to the length of data window used. On the other hand the dispersion of the average solution from the group of sessions decreases with growing number of sessions in the group investigated. The plots illustrating σ_z and σ_m dispersions for vertical component of the vector calculated with the Bernese software using 3h sessions with 0.5h shift over 3 months are given in Fig. 2a and Fig. 2b, respectively.

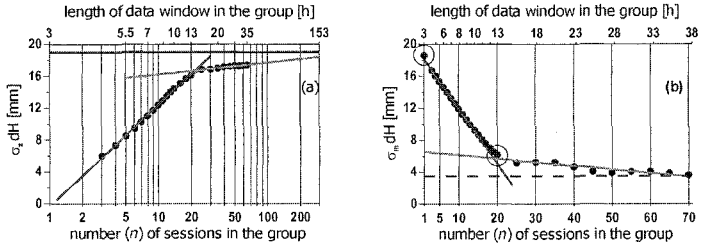


Fig. 2. Mean st. dev. σ_z in the group of n consecutive solutions for vertical component of the vector (a) and mean st. dev. of the mean of n consecutive solutions in the group (b). Grey line in (a) corresponds to the st. dev. in 3 months long group of sessions. Dashed line in (b) corresponds to the accuracy of average solution from 3 months data

The change of the variation rate of dispersion (Fig. 2 a and b) is observed around data window of 12h. For longer windows the variation rate of dispersion becomes substantially less significant. The mechanisms that affect GPS solutions obtained from sessions shorter than 12h differ from those observed in the solutions from longer sessions. Solutions from sessions shorter than 12h are mainly affected with noise and periodic biases due to GPS satellite constellation. Fig. 2b also shows that vertical component of 42 km vector can be determined from 12h GPS data with accuracy of 6-7 mm.

The increase of the length of session (data window) used to generate GPS solutions, results in reduction of dispersion of those solutions, mainly due to averaging both noise and periodic biases. Time series of GPS solutions based on longer sessions is much smoother as compared with the one derived from short sessions. The effect of such a smoothing procedure could be simulated with combining and averaging GPS solutions obtained using short data window. The effect of smoothing GPS-derived vertical component of the vector obtained from 3h sessions, by applying running average with 12.5h window may be seen when comparing Fig. 3a with Fig. 3b.

The results shown so far indicate the external accuracy estimate of GPS solutions that is based on analysis of repeatability with use of time series of high temporal resolution. Such an accuracy estimate does not coincide with standard deviations provided by GPS processing software that reflect an internal

accuracy of the system. Short period biases, including non-modelled effects, some with unstable amplitudes, that affect GPS solutions are as systematic-type terms not reflected in uncertainty estimation. Therefore, calculated uncertainty is usually much too optimistic as the accuracy estimate. Comparison of external with internal accuracy of determination of the length and vertical component of the vector investigated using the Bernese and Pinnacle software from sessions of different length with 1h shift, is given in Fig. 4 and Fig. 5, respectively.

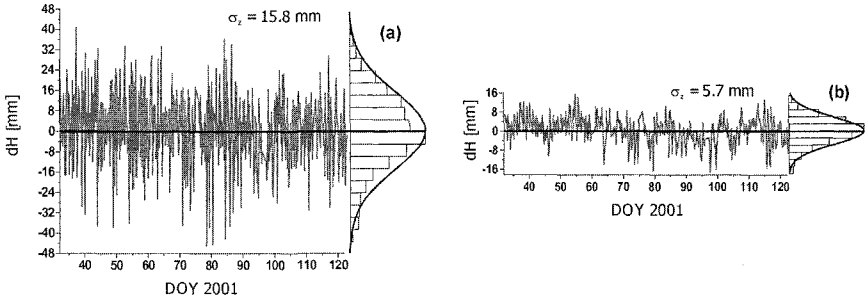


Fig. 3. Time series of vertical component of the vector obtained from processing 3h sessions with 30m shift (a), and running average with a window of 12.5h (averaging the current groups of 20 solutions from single sessions) (b)

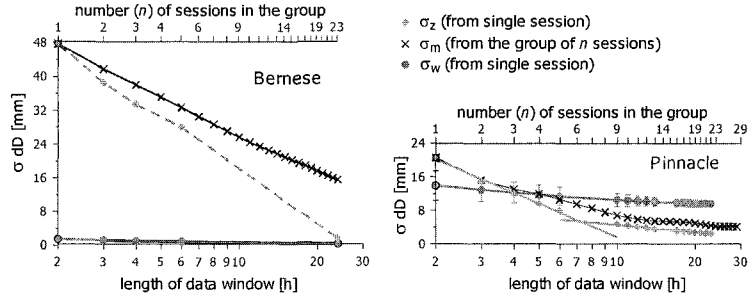


Fig. 4. St. dev. σ of the GPS-derived vector lengths provided by processing softwares, and estimated by statistical analysis of time series of GPS solutions

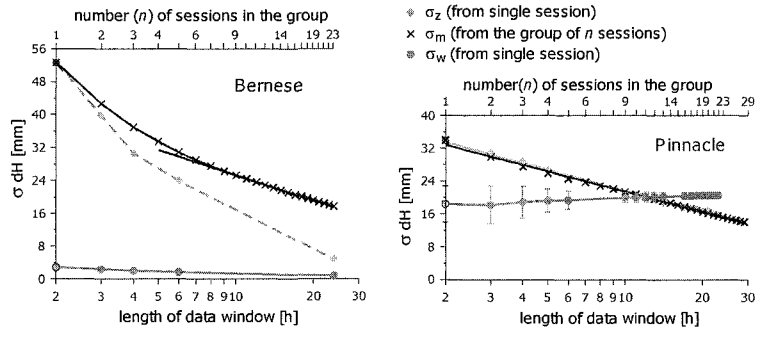


Fig. 5. St. dev. σ of the GPS-derived vertical component provided by processing softwares, and estimated by statistical analysis of time series of GPS solutions

The dispersion of GPS solutions obtained using the Bernese software is larger than that from the Pinnacle software, in particular when short sessions were processed (Fig. 4 and Fig. 5). That phenomenon is more distinct for vector length than for the vertical component. In spite of large dispersion of GPS solutions that might be explained by the use of QIF strategy for processing data for 42 km vector with the Bernese software, the software-provided standard deviations of the solutions obtained are at the very low level. Both vertical component and the length of a vector examined were determined with the accuracy of about 5 cm from 2h sessions while precision of the solution provided by the Bernese software was at the level of single millimetres. In case of solutions based on 24h sessions the discrepancy between the external and internal estimate of accuracy drops down to the level of a few millimetres holding their ratio at the same level.

The effect of noise and periodic biases on GPS solutions can also be reduced by averaging the solutions over the groups of sessions, e.g. the mean of *n* 2h sessions, that form the group. Such simple

averaging does not remove all effects that are eliminated when processing with the Bernese software one session of length corresponding to the length of the respective group of short sessions. The external accuracy based on analysis of groups of sessions is thus overestimated although its trend remains similar to the one related to single sessions.

For GPS solutions obtained using the Pinnacle software being less sophisticated in terms of GPS observations modelling than the Bernese one, the main trends for the external accuracy getting improved with growing session length are preserved. Different image has, however, the mutual relationship of the external and internal accuracy. The internal accuracy estimate given by Pinnacle is more realistic than in case of the Bernese software. Vertical component of the vector examined was determined from 2h sessions with the Pinnacle software with the accuracy of about 3.5 cm while software-provided precision of the solution was about 2 cm. For the length of the vector investigated the external and internal accuracy was 2 cm and 1.5 cm, respectively. Moreover, for a certain length of session, internal accuracy coincides with the external one. With further growing session length the internal accuracy of the solutions becomes overestimated. Such a singularity corresponds to 12h and 4h sessions in case of vertical component and vector length determination, respectively. GPS solutions from the Pinnacle software, averaged over the groups of sessions, practically coincide with those corresponding to the respective single sessions. It particularly concerns vertical component.

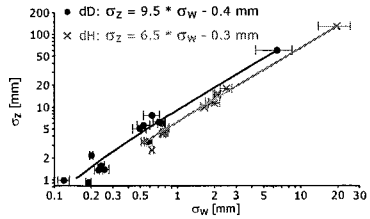


Fig. 6. Correlation between the external accuracy σ_z and the internal accuracy σ_w of GPS solution provided by the Bernese software

The discrepancy between the external accuracy σ_z and internal consistency σ_w of GPS solutions obtained using the Bernese software was investigated for numerous vectors of length from a few tens to a few thousand kilometres. 3h and 24h GPS sessions covering 4 months of 2001 were processed. The correlation between the external and the internal accuracy estimates is given in Fig. 6. Internal accuracy differs from the external one by a scale factor of about 7 and 10 for a vertical component and vector length, respectively.

3. Strategy of GPS solution quality analysis

The strategy developed for detecting and modelling biases in time series of GPS solutions is given in Fig. 7.

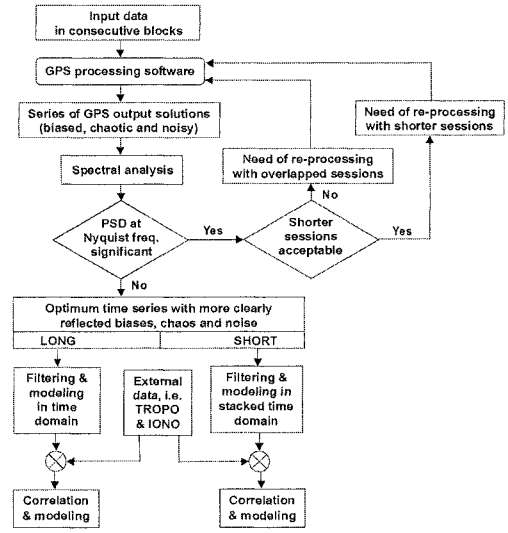


Fig. 7. Flowchart of the strategy of GPS solutions quality analysis

Temporal resolution of a series of GPS solutions is determined by a sampling rate that corresponds to the length of session when data is processed in consecutive blocks. The longer the processed GPS sessions the smoother become solutions and consequently time series obtained. Smoothing obviously reduces random effects but also some periodic biases. Solutions based on shorter sessions are thus affected by larger biases. To study biases in GPS solutions the examination of time series with sufficient temporal resolution is required. Thus time series of GPS solutions obtained from short sessions is preferable despite of increased noise level. Shortening the session is, however, limited by the length of the vector determined. To increase temporal resolution of time series the overlapped sessions need to be processed. Overlapping causes an increase of correlation between consecutive GPS solutions that has to be carefully considered when estimating statistical parameters of such time series. It allows, however, for efficient detection and separation of chaotic effects from biases.

Power spectrum density at the Nyquist frequency can be used as an indicator of need of a

further increase of temporal resolution of a series by processing either shortened sessions or more overlapped sessions. Time series, optimal with respect to the structure of investigated biases, consist of GPS solutions obtained from optimum session length and optimum sampling time. Such a series reflects clearly periodic biases as well as chaotic terms. It is therefore suitable to detect and estimate periodic biases. Similar strategy with 4h window stepped by 30 minutes and followed by a running average procedure was used for investigation of the vertical shift caused by Earth tides (Neumeyer et al., 2002).

Long time series can directly be processed for filtering noise and modelling periodic biases. In case of short time series stacked solutions need to be calculated over e.g. one-day period and then filtered and processed for modelling biases. Finally the models could be correlated with the external data, e.g. troposphere or ionosphere parameters to separate biases and find their sources.

Conclusions

A dispersion of GPS-derived vector components considered as an external accuracy estimate does not coincide with processing software-provided estimated accuracy that is the internal accuracy estimate. The discrepancy between externally and internally estimated accuracy was investigated using statistical analysis of time series of vector components obtained with the Bernese and Pinnacle software. Internal accuracy provided by the Bernese software differs from the external one, in cases investigated, by a scale factor of about 7 and 10 for a vertical component and vector length, respectively. Internal accuracy estimation provided by the Pinnacle software can be considered as the acceptable rough estimate of accuracy.

Accuracy of GPS solutions based on data from permanent stations or from long-term geodynamics campaigns can efficiently be estimated by investigating time series of overlapping solutions using the tools of statistical analysis. The described strategy of quality analysis of GPS solutions besides filtering allows for estimation of biases and chaotic effects. That procedure is not suitable for estimation of accuracy of GPS solutions obtained from single, short time occupation of sites. The majority of noise can be filtered using simplified statistical analysis of overlapped solutions based on sub-intervals of the observed session. Biases and jerks, however, can only be roughly estimated externally using the results of extended statistical analyses.

GPS positioning based on shorter than 12h single observing session could not provide the solution for vector components with accuracy below 1 cm, even for short vectors.

Acknowledgements

The research was partially financed by the Polish State Committee for Scientific Research (grant PBZ-KBN-081/T12/2002). The authors express special gratitude to Mrs. M. Mank and Mr. L. Zak from the Institute of Geodesy and Cartography, Warsaw, as well as to Dr. P. Wielgosz from the University of Warmia and Mazury, Olsztyn, for processing GPS data.

References

Bruyninx C., (2001): Overview of the EUREF Permanent Network and the Network Coordination Activities, EUREF Publ., Eds. J. Torres, H. Hornik, *Bayer. Akad. der Wies.*, München, Germany, 2001, No 9, pp. 24-30.

Gandolfi S., Gusella L., Perfetti N., Dubbini M., (2003): Accuracy and precision vs software and different conditions using Italian GPS fiducial network (IGFN) data, *Reports on Geodesy*, WUT, Warsaw, No 2(65) 2003, pp 65-71.

Harris F.J., (1978): On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform, *Proc. IEEE*, Vol. 66, January 1978, pp. 51-83.

King M., Coleman R., Nguyen L.N., (2003): Spurious Periodic Horizontal Signals in Sub-Daily GPS Position Estimates, *J. of Geod.*, Vol. 77, Nr 1-2, May 2003, pp. 15-21.

Krynski J., Zanimonskiy Y.M., Wielgosz P., (2000): Short Time Series Analysis of Precise GPS Positioning, Presented at the *XXV Gen. Assembly of EGS*, Nice, France, 25-29 April 2000.

Krynski J., Cisak J., Zanimonskiy Y.M., Wielgosz P., (2002a): Variations of GPS Solutions for Positions of Permanent Stations – Reality or Artefact, EUREF Publ. No 12, *Mitteilungen des Bundesamtes für Kart. und Geodäsie*, Frankfurt am Main, Band 29, pp. 320-325.

Krynski J., Zanimonskiy Y.M., Wielgosz P., (2002b): Modelling biases in GPS Positioning Based on Short Observing Sessions, Presented at the *XXVII Gen. Assembly of EGS* 2002, Nice, France, 21-26 April 2002.

Krynski J., Zanimonskiy Y.M., (2002): Quality Structure of Time Series of GPS Solutions, *Proc. of the NKG Gen. Assembly*, 1-5 Sept. 2002, Espoo, Finland, pp. 259-264.

Neumeyer J., Barthelmes F., Combrink L., Dierks O., Fourie P., (2002): Analysis of Results from the SG Registration with the Dual Sphere Supercon-ducting Gravimeter at SAGOS (South Africa), BIM 135, 15 July 2002, Paris, pp. 10607-10616.

Poutanen M., Koivula H., Ollikainen M., (2001): On the Periodicity of GPS Time Series, *Proc. of IAG 2001 Sc. Assembly*, 2-7 Sept. 2001, Budapest, Hungary.

Teunissen P.J.G., (2002): The Parameter Distributions of the Integer GPS Model, *J. of Geod.*, Vol. 76, Nr 1, January 2002, pp. 41-48.

Realistic Uncertainty Measures for GPS Observations

S. Schön and H. Kutterer
Deutsches Geodätisches Forschungsinstitut, Marstallplatz 8, D-80539 Munich, Germany

Abstract. The classical concept in geodesy of an exclusively stochastic assessment of the total error budget of observation data is extended. Uncertainty due to remaining systematics (imprecision) is treated consistently by interval mathematics. The superposition of both random variability (stochasticity) and imprecision yields uncertainty measures such as extended point confidence regions which are more realistic. The new concept is applied to GPS phase measurements. It is exemplarily discussed for a synthetic GPS network.

Keywords. Systematic Errors, Imprecision, Interval Mathematics, GPS

1 Introduction

It is common practice in geodesy to consider all types of uncertainty in geodetic observations in the end as random (stochasticity). Typical measures of stochasticity are confidence regions such as point confidence ellipsoids. This is consistent with the recommendations of the Guide to the Expression of Uncertainty in Measurement (GUM), ISO (1995).

However, due to remaining systematics there is a second type of uncertainty of its own right (imprecision). It is caused by intentional model simplifications or imperfect knowledge of external influences on the observations. Imprecision can be treated by means of interval mathematics, cf. Schön and Kutterer (2001a,b). In this case, interval boxes and convex polyhedra (zonotopes) are meaningful uncertainty measures for point positions.

In the following, the use of intervals is shown for the modeling of imprecision in GPS phase observations. The interval extension of the least-squares estimates transfers the imprecision from the observations to the estimated parameters. Exemplary results for a synthetic GPS network show the different reaction of stochasticity and imprecision measures on an increased number of observations. Finally, confidence ellipsoids and zonotopes are superposed to extended uncertainty measures of point positions.

2 Remaining systematics in GPS phase observations

The geodetic model for the estimation of GPS point positions is based on the Euclidean distance between the satellite S^k and the receiver P_i. Different corrections are necessary in order to take effects into account like tropospheric or ionospheric refraction, or eccentricities caused by instrumental heights as well as phase center variation or offsets, cf. Teunissen and Kleusberg (1998).

The correction steps for GPS observations are rather sophisticated. They depend mainly on the size and purpose of the network. For this reason different analysis strategies exist. Some effects may be partially corrected, but residual terms (like tropospheric delays) can also be estimated as additional parameters. Moreover, the impact of perturbing effects can be reduced or eliminated by various ways of observation differencing (zero, single, double, or triple differences) without explicit modeling.

The basic relation for the carrier phase measurements (in $[m]$) between the station P_i and the satellite S^k at epoch t expressed in GPS time reads as:

$$\begin{aligned}\Phi_i^k(t) \;=\; & \varrho_i^k(t) + I_i^k(t) + T_i^k(t) + m_i^k(t) + \\ & c\,(\delta_i(t) - \delta^k(t - \tau_i^k)) + \\ & c\,(dt_i(t) - dt^k(t - \tau_i^k)) + \\ & \lambda\,(\phi_i(t) - \phi^k(t - \tau_i^k)) + \lambda\, N_i^k. \end{aligned} \tag{1}$$

The first term of Eq. (1) is the basic geometric model. It equals the Euclidian distance between the satellite and the receiver at epoch t. The signal travel time is given by τ_i^k. Corrections I_i^k and T_i^k are applied due to ionospheric and tropospheric refraction, respectively. The fourth term indicates the impact of signal multipath. The fifth term contains the clock errors terms, and the sixth term is introduced to handle delays in the measurement equipment. The last two terms result from non-zero initial phases and the carrier phase ambiguity, cf. Teunissen and Kleusberg (1998). In general, a noise term $\varepsilon_i^k(t)$ is associated with the phase measurement which describes the stochasticity of the observations.

Despite all correction steps some systematic effects remain after the data analysis due to only approximately or imprecisely known external influences. In the following an interval-based strategy is proposed to tackle this problem. The starting point is to derive interval radii as measures of imprecision.

The impact of the correction for ionospheric refraction is discussed exemplarily. It can be expressed in a first order approximation by:

$$I_i^k = \frac{\alpha \cdot TEC}{f^2 \cdot cos\zeta_i^k}, \tag{2}$$

where $\alpha = -40.30$ denotes a constant, TEC the total vertical electron content, f the carrier frequency, and ζ_i^k the zenith angle to the satellite.

First, all influence parameters $\mathbf{p}$ which are only imprecisely known have to be identified. Here, these are the parameters $\alpha, TEC, f, \zeta_i^k$. They can be modeled by intervals as $[p_m - p_r, p_m + p_r]$. The actual values of the parameters correspond to the interval midpoints p_m and the imprecision to the radii p_r.

The magnitude of the imprecision can be derived for example from the temporal range of parameters like $[TEC] = [0.99, 1.01] \cdot 10^{17} \frac{e}{m^2}$ what is based on Klobuchar (1996) and the assumption of a 1% relative uncertainty. The sensor performance like the stability of the frequency given by $\frac{df}{f} = 10^{-12}$ yields $[f] = f + [-10^{-12}, 10^{-12}]\, f$.

Considering orbit errors or imprecisely known station coordinates of 1 m in each component as well as discrepancies between the Euclidian distance and the actual electromagnetic propagation way, an imprecision of $d\zeta_i^k = 0.1''$ could be assigned to the zenith distance. Model constants are only given with a limited number of significant digits, hence $[\alpha] = [-40.305, -40.295]$.

Model approximations or unmodeled effects could also be enclosed by intervals. E.g., Eq. (2) does not take higher order terms or effects due to ray bending into account, cf. Brunner and Gu (1991). Other correction formulas can be treated in a similar way.

The main goal of the interval-based strategy is to quantify for maximum impact of the influence parameters on the corrected observations (worst case). The sensitivity analysis of correction models and analysis strategy is a straightforward approach for this problem. The total differential of Eq. (1) w.r.t. to the influence parameters at a given epoch t yields:

$$\begin{aligned} d\Phi_i^k \;=\; & d\varrho_i^k(\mathbf{x}_i, \mathbf{x}^k) + dI_i^k(\alpha, TEC, f, \zeta_i^k) + \\ & dT_i^k(\vartheta_i, p_i, e_i, c_j, \zeta_i^k \ldots) + \ldots + \Delta_i^k. \end{aligned} \tag{3}$$

For simplicity, only the first three terms and an additional one are given. The first one denotes the total differential of the pseudorange with respect to the station coordinates $\mathbf{x}_i$ and the satellite coordinates $\mathbf{x}^k$. The second term describes the impact of the ionospheric refraction model. The third one denotes the total differential of the tropospheric refraction model.

These models are in general influenced by the meteorological parameters like, e.g., temperature ϑ_i, pressure p_i, or humidity e_i, by model constants c_j and by the zenith angle ζ_i^k of the satellite. The other terms of Eq. (1) can be differentiated in the same way. The last term Δ_i^k quantifies unmodeled additional effects and rounding-off errors.

With $\mathbf{f}_i^k$ the vector of partial derivatives and $\mathbf{dp}$ the vector of differentials of the above-mentioned influence parameters, Eq. (3) can be rewritten as:

$$d\Phi_i^k = (\mathbf{f}_i^k)^T\, \mathbf{dp}. \tag{4}$$

The maximum values of the differentials $\mathbf{dp}$ are considered as imprecision measures (interval radii) of the corresponding influence parameters:

$$\begin{aligned} \Phi_{i\,r}^k &:= |(\mathbf{f}_i^k)^T|\, \mathbf{p_r}, \\ \Phi_{i\,m}^k &:= \mathbf{f}_i^k\, \mathbf{p_m} = \Phi_i^k. \end{aligned} \tag{5}$$

Finally, the intervals of the corrected observations which describe adequately the imprecision are reconstituted with the interval midpoint and radius:

$$[\Phi_i^k] = [\, \Phi_{i\,m}^k - \Phi_{i\,r}^k\,,\; \Phi_{i\,m}^k + \Phi_{i\,r}^k\,]. \tag{6}$$

The following two figures indicate the impact of the imprecision of different parts of the correction model acc. to Eq. (1) on the corrected phase observations. Fig. (1) presents the interval radii for the original observations (zero differences, ZD), Fig. (2) the values which result for the corresponding double differences (DD). The values are derived from a typical GPS short-range configuration (baseline lengths of several km). For the tropospheric refraction the Saastamoinen (1973) formula is used. Note the different scales of the up-axes in the two figures.

The different contributions to the interval radii are divided into four groups. The satellite-specific part is caused by the imprecision of the orbit and the frequency or clock stability. The propagation-specific part assesses remaining systematics of the correction models. This contribution depends strongly on the elevation. Finally, eccentricities of the station coordinates as well as receiver clock stability are considered as station-specific. A usually unmodeled part of $\Delta_i^k = 0.1\sigma_\Phi$ is also taken into account.

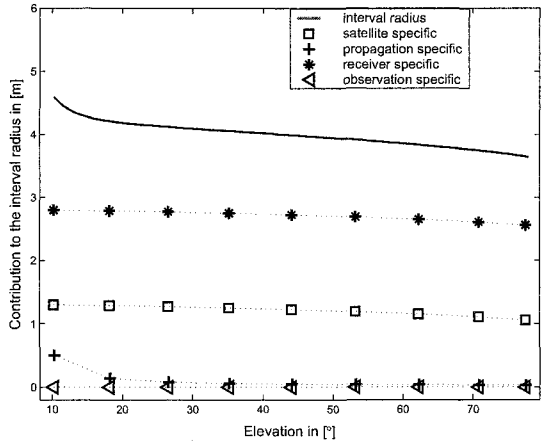


Fig. 1. Contributions to the interval radius of GPS phase measurements (zero differences)

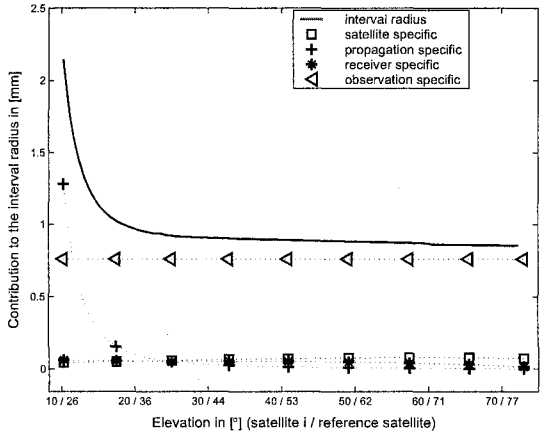


Fig. 2. Contributions to the interval radius of GPS phase measurements (double differences)

This approach is useful to analyse the contributions of particular influence factors to the whole imprecision. Similar studies were carried out for EDM measurements which are presented in detail in Schön and Kutterer (2001a,b).

Considering the station P_i and the reference station P_j, the satellite S^k and the reference satellite S^l, the DD read as:

$$\Phi_{i,j}^{k,l} = (\Phi_i^k - \Phi_j^k) - (\Phi_i^l - \Phi_j^l) =: \mathbf{M}\Phi. \quad (7)$$

When all transposed vectors $(\mathbf{f}_i^k)^T$ are compiled in a matrix $\mathbf{F}$ and Eq. (5) is applied to Eq. (7), the interval midpoint and radius of the DDs are given by:

$$\begin{array}{lcl} \Phi_{i,j\,m}^{k,l} & := & \mathbf{M}\,\mathbf{F}\,\mathbf{p_m} = \Phi_{i,j}^{k,l}, \\ \Phi_{i,j\,r}^{k,l} & := & |\mathbf{M}\,\mathbf{F}|\,\mathbf{p_r}. \end{array} \quad (8)$$

3 Interval extension of least-squares adjustments

The intervals which describe the corrected DD GPS phase measurements are now propagated to the estimated coordinates. It is assumed that the ambiguities are resolved and no additional parameters like residual tropospheric delays are modeled. The following is based on the interval extension of the least-squares estimator as described by Schön and Kutterer (2001a,b). Stochasticity and imprecision are treated separately.

Let $E(\mathbf{l})$ be the expectation of the $n \times 1$ vector of observations $\mathbf{l}$, $\mathbf{A}$ the $n \times u$ matrix of the partial derivatives, and $\mathbf{a_0}$ the zero-order vector of the Taylor expansion. Then a linearized Gauss-Markov model, cf. Koch (1999), is given by:

$$E(\mathbf{l}) = \mathbf{A}\mathbf{d}\hat{\mathbf{x}} + \mathbf{a_0}, \quad D(\mathbf{l}) = \sigma_0^2 \mathbf{P}_{ll}^{-1}. \quad (9)$$

with $D(\mathbf{l})$ the dispersion of the observations, $\mathbf{P}_{ll}$ their regular weight matrix and σ_0^2 the a-priori variance factor. Assuming uncorrelated original phase observations (zero differences, ZD), the dispersion matrix of the DDs yields

$$D(\mathbf{l}) = \sigma_0^2\, \mathbf{M}\mathbf{M}^T, \quad (10)$$

where $\mathbf{M}$ indicates the matrix of the differencing method, cf. Eq. (7). A no-net-translation constraint is added in order to solve the datum problem:

$$\mathbf{d}\hat{\mathbf{x}} = \mathbf{Q}_{\hat{x}\hat{x}}\mathbf{A}^T\mathbf{P}_{ll}(\mathbf{l} - \mathbf{a_0}), \quad \mathbf{Q}_{\hat{x}\hat{x}} = (\mathbf{A}^T\mathbf{P}_{ll}\mathbf{A})^-. \quad (11)$$

An interval extension of Eq. (11) which is obtained by replacing the vector of the corrected observations $\mathbf{l}$ by the interval vector $[\mathbf{l}]$ reads as

$$[\mathbf{d}\hat{\mathbf{x}}] = (\mathbf{Q}_{\hat{x}\hat{x}}\mathbf{A}^T\mathbf{P}_{ll})\,([\mathbf{l}] - [\mathbf{a_0}]). \quad (12)$$

Eq. (12) can be split into midpoint and radius. The midpoint $\mathbf{x_m}$ is identical with the classical least-squares estimate. The imprecision is linearly propagated from the observations to the estimates. As the real intervals are not closed w.r.t. to a linear mapping according to Eq. (11) the zonotope

$$\mathcal{Z} := \{\mathbf{d}\hat{\mathbf{x}} \in R^u \,|\, \mathbf{d}\hat{\mathbf{x}} = \mathbf{Q}_{\hat{x}\hat{x}}\mathbf{A}^T\mathbf{P}_{ll}(\mathbf{l} - \mathbf{a_0}), \mathbf{l} \in [\mathbf{l}]\} \quad (13)$$

is the correct description of the imprecision of the point position. Its tightest interval inclusion is given by the interval box which is centered in $\mathbf{x_m}$ and spanned by the radius vector $\mathbf{x_r}$, cf. Eq. (12).

4 Example: GPS Network

In this section the characteristics of stochasticity and imprecision are discussed for a simulated GPS network. Fig. (3) shows the sky distribution of the satel-

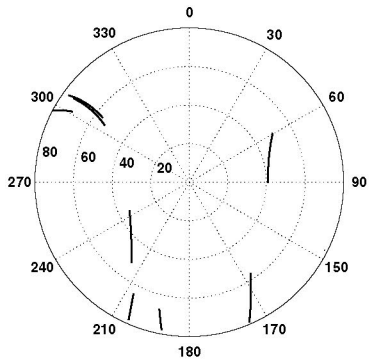


Fig. 3. Satellite sky distribution

lites during one hour of observation time. The simulated local network consists of eight points located approximately at 49° N Lat. and 9° E Long. Seven baselines are selected with lengths from 1 km to 14 km and with the first point as reference point. The standard deviation of the ZDs is $\sigma_\Phi = 2\ mm$ (L1 observations only).

A standard double differencing approach is used in order to reduce the impact of the major effects. The resulting interval radii $\Phi^{k,l}_{i,j\ r}$ after DDing vary between 0.8 and $2.1 mm$ depending on the zenith distance differences of the considered satellite-station pairs. A cut-off angle of 10° is used. For a processing interval of $\Delta t_1 = 1200\ sec$, $n_1 = 196$ DDs can be computed (Scenario I).

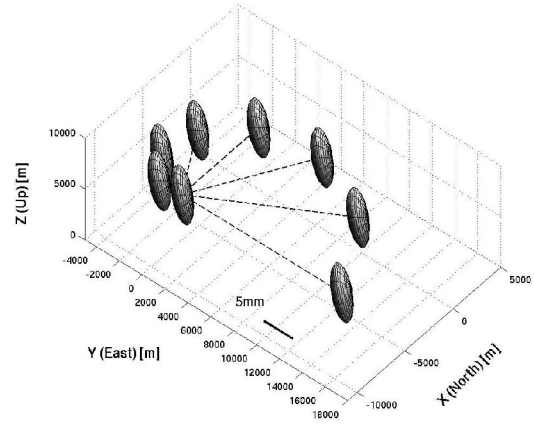


Fig. 4. 3d 95% confi dence ellipsoids

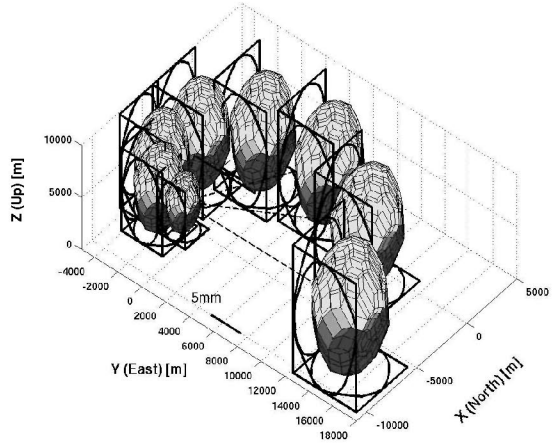


Fig. 5. 3d zonotopes

Fig. (4) shows the 95% confidence ellipsoids of the estimated point positions resulting from the simulation. These ellipsoids are given in a unique local topocentric coordinate system for all stations. The lengths of the semi-axes are $a = 5.4\ mm, b = 2.1\ mm$, and $c = 1.3\ mm$. The standard deviations of the coordinates are equal for all points: $\sigma_N = 0.4\ mm, \sigma_E = 0.5\ mm$, and $\sigma_{Up} = 3.7\ mm$. The semimajor axis of the 95% confidence ellipsoids is oriented to the South-East ($Az = 140°, El = 82°$). This is due to the specific satellite distribution depicted in Fig. (3) which shows some gaps in the North-South direction.

Fig. (5) presents the corresponding three-dimensional zonotopes (grey-shaped) which describe the imprecision of the estimated point positions. In addition, the enclosing interval boxes are plotted. Interval boxes and zonotopes are computed with respect to Eqs. (12) and (13). The interval radii of the coordinates vary from 2.3 to 5.1 mm (x_r), 2.3 to 5.2 mm (y_r), and 5 to 12.3 mm (z_r). The orientations of the zonotopes are similar to those of the confidence ellipsoids. This is due to the geometry of the scenario contained in the matrix $\mathbf{A}$. For the confidence ellipsoids the concrete shape is equal for all network points. Obviously, the choice of the reference point and reference satellite does not affect the uncertainty of the estimated point positions.

At present, this is in contrast to the zonotopes where the propagation of the interval radii is split into two steps. First, the radii of the sensor parameters are transferred to the ones of the DD observations. Then, these radii are propagated to the parameters. A consistent one-step procedure is currently in preparation.

Thanks to the inclusion property of interval mathematics this is not critical as the range of values is a subset of the resulting zonotopes.

By means of a second simulation (senario II) with a processing rate of $\Delta t_2 = 360\ sec$ for the same observation time the influence of the number of observations on the above presented uncertainty measures was assessed. Thus, only the number of observations was increased to $n_2 = 518$ double differences. Below, the two-dimensional projections of the three-dimensional 95% confidence ellipsoids and zonotopes are presented. Fig. (6) gives a represen-

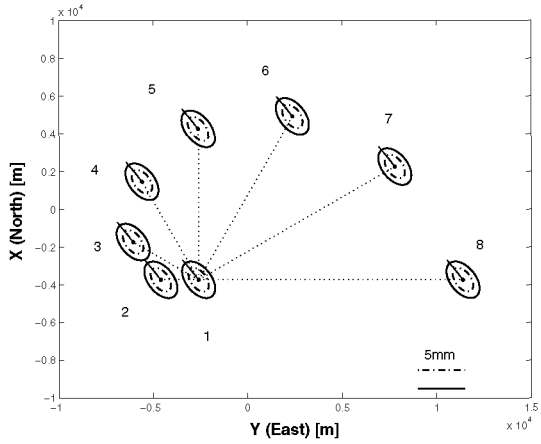


Fig. 6. Comparison of projected 3d 95% confi dence ellipsoids

tation of the projected three dimensional confidence ellipsoids for both scenarios with scenario I in solid grey and scenario II in dash-dotted black. Due to the quadratic variance-covariance propagation which is intrinsic to the classical stochastic approach, all confidence ellipsoids of scenario II are enclosed by those of scenario I because of more observations. The standard deviations yield $\sigma_N = 0.1\ mm, \sigma_E = 0.2\ mm$, and $\sigma_{Up} = 1.5\ mm$. The reduction w.r.t. scenario I is 60% on average.

Fig. (7) shows a corresponding comparison of the projected zonotopes. For these convex polytopes a totally different behavior is shown. The results of scenario I are plotted in grey, those corresponding to scenario II in black (dash-dotted line). Contrary to the confidence ellipsoids, the zonotopes of scenario II have about the same size as those of scenario I: 2.3 to 4.9 mm (x_r), 2.3 to 5.1 mm (y_r), and 5.5 to 12.2 mm (z_r).

Two counteracting mechanisms play a key role in the propagation of imprecision. On the one hand, additional observations strengthen the geometry and yield a more favourable propagation of interval radii

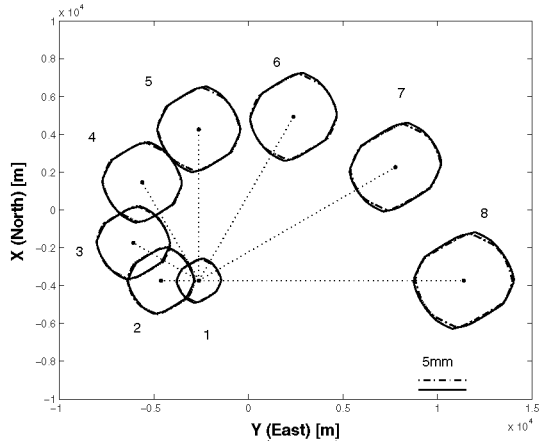


Fig. 7. Comparison of projected 3d zonotopes

like for the variances. On the other hand, the total uncertainty budget increases. Here, the influence of both mechanisms is balanced as the resulting interval radii decrease at maximum by 5% (height component). Schön and Kutterer (2001b) obtained similar results for a two-dimensional network. Moreover, an improved network geometry can partially reduce the imprecision.

5 Combination of stochasticity and imprecision

Stochasticity and imprecision measures of point positions can be superposed yielding the *extended uncertainty*. As the main idea, confidence regions and zonotopes are considered as convex sets. Let $\mathcal{E}_{1-\alpha}$ denote the confidence ellipsoid of the confidence level α and $\mathcal{Z}$ the zonotope. Then both convex sets are combined using the *Minkowski-sum*.

$$\mathcal{C}_E := \mathcal{E}_{1-\alpha} \oplus \mathcal{Z} = \{\mathbf{y} + \mathbf{z} \mid \mathbf{y} \in \mathcal{E}_{1-\alpha}, \mathbf{z} \in \mathcal{Z}\}, \tag{14}$$

where $\mathcal{C}_E$ denotes the *extended point confidence region*. Taking the imprecision of the estimated point position into account, the extended confidence region covers the true position with a probability of $1 - \alpha$.

Fig. (8) shows the construction principle for the simple case of a zonotope with 6 vertices. Geometrically the zonotope is decomposed in sectors corresponding to its complete fan, cf. Henk et al. (1997). The resulting sectors are assigned to the corresponding points of the ellipse (direction of the ellipse tangent equals the direction of the zonotope edge).

Fig. (9) depicts the extended point confidence re-

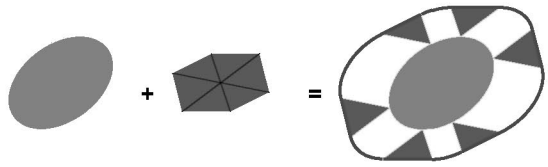


Fig. 8. Construction principle of the extended uncertainty region

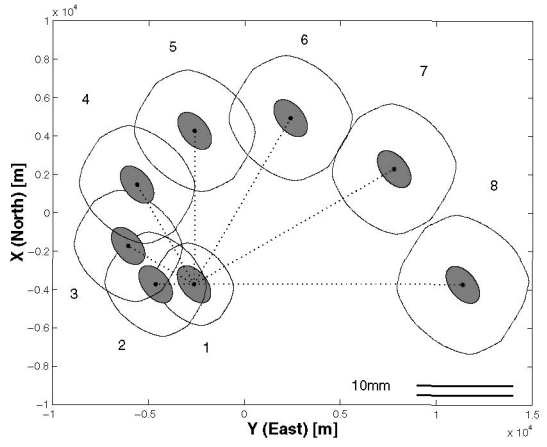


Fig. 9. Extended uncertainty regions for the GPS network of scenario I (stochasticity+imprecision)

gions for the above-discussed scenario I of the GPS network. Similar results are given by Kutterer (2002) for the extension of GPS phase ambiguity search spaces. He used an optimization approach to combine stochasticity and rated imprecision (modeled by fuzzy sets).

6 Conclusions

In this paper an extended approach was presented for the joint treatment of the two most relevant types of uncertainty in GPS data analysis, namely stochasticity and imprecision. In many geodetic applications both types have to be taken into account. Hence, the classical stochastic description of uncertainty is incomplete and has to be extended. The effects of imprecision are adequately treated by interval mathematics. It was outlined how meaningful observation intervals can be derived in geodetic GPS practice. Using the interval extension of the least-squares estimator, this imprecision can be propagated to the estimated parameters yielding zonotopes in general.

Finally, the application of the extended approach to a synthetic short-range GPS network was shown. Obviously, the extended uncertainty measures derived by superposition of confidence regions and zonotopes are only partially affected by the actual number of observations. Hence, they are reflecting the situations in geodetic practice quite well as they can not decrease beyond any limit. It can be stated that the corresponding interval-mathematical problems are mostly solved. However, further efforts are necessary in order to derive practicable interval radii for all relevant geodetic applications of the GPS.

Acknowledgment. The presented paper shows results and new ideas developed during the research project KU 1250/1-1/2 "Optimaler Entwurf geod¨atischer Überwachungsnetze unter Ber¨ucksichtigung strenger Toleranzen", which is sponsored by the Deutsche Forschungsgemeinschaft (DFG). This is gratefully acknowledged by the authors. Thanks to Klaus Kaniuth and Manuela Kr¨ugel (both at DGFI) for helpful advice on the Bernese Software.

7 References

Brunner F. and M. Gu (1991): An improved model for the dual frequemcy ionospheric correction of GPS observations. *manuscripta geodaetica*, 16:205-214.

Henk M., J. Richter-Gebert and G. Ziegler (1997): Basic Properties of convex Polytopes. In: J. Goodman and J. O'Rourke, editors. *Handbook of Discrete and Computational Geometry*, CRC Press, Boca Raton, pp.243-270.

ISO (1995): *Guide to the Expression of Uncertainty in Measurements*. International Organization of Standardization.

Koch K.-R. (1999): *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer Berlin.

Klobuchar J. (1996): Ionospheric effects on GPS. In: B. Parkinson and J. Spilker, editors. *Global Positioning System: Theory and Applications*, American Society of Aeronautics and Astronautics, Vol.163:485-502.

Kutterer H. (2002): Joint treatment of random variability and imprecision in GPS data analysis. *Journal of Global Positioning Systems*, 1(2):96-105.

Saastamoinen J. (1973): Contribution to the theory of atmospheric refraction. *Bulletin Géodésique*, 107:13-34.

Sch¨on S. and H. Kutterer (2001a): Interval-based Description of Measurement Uncertainties and Network Optimization. In: A. Carosio and H. Kutterer, editors. *Proceedings of the First International Symposium on Robust Statistics and Fuzzy Techniques in Geodesy and GIS*. Report No.295, Swiss Federal Institute of Technology Zurich (ETH), Institute of Geodesy and Photogrammetry, pp.41-46.

Sch¨on S. and H. Kutterer (2001b): Optimal Design of Geodetic Monitoring Networks by means of Interval Mathematics. *Proceedings of the* 10^{th} *FIG International Symposium on Deformation Measurements held in Anaheim California, March 19-22 2001*, pp.362-371.

Teunissen P. and A. Kleusberg (1998): GPS Observation Equations and Positioning Concepts. In: P. Teunissen and A. Kleusberg (editors). *GPS for Geodesy*, 2^{nd} ed.. Springer Berlin, pp.187-230.

Networked Transport of RTCM via Internet Protocol (NTRIP)

G. Weber, D. Dettmering
Bundesamt für Kartographie und Geodäsie (BKG),
Richard-Strauss-Alle 11, 60598 Frankfurt am Main, Germany

H. Gebhard,
Lehrstuhl für Kommunikationstechnik, Universität Dortmund,
Otto-Hahn-Str. 4, 44221 Dortmund, Germany

Abstract. Within the framework of EUREF, the Federal Agency for Cartography and Geodesy (BKG) has developed a new technique using the open Internet for the real-time collection and exchange of GNSS data, as well as for broadcasting derived products. A major purpose of these activities is the dissemination of Differential GPS corrections (DGPS) for precise positioning and navigation. This paper describes the http-based technique for streaming GNSS data to mobile users connected to the Internet via Mobile IP-Networks like GSM, GPRS, EDGE, or UMTS. The technique establishes a format called Networked Transport of RTCM via Internet Protocol (Ntrip).

Keywords: Differential GPS, Internet, Real-Time, Positioning, Navigation

1 Introduction

Due to the increased capacity of the Internet, applications which transfer continuous data-streams by IP-packages (IP for Internet Protocol), such as Internet Radio, have become well-established services. Compared to these applications, the bandwidth required for the transfer of real-time GPS positioning data is relatively small. As a consequence, the dissemination of DGPS corrections via Internet and mobile IP-Networks becomes an alternative to the usage of other terrestrial broadcasting techniques. Wireless Internet access is available nowadays in many regions. No significant degradation of positioning performance is expected when using it, but some remarkable advantages are possible. In addition, data streams from reference stations and (GIS-) data bases can be accessed simultaneously through one communication technique. Stationary applications may use a wired Internet connection through a desktop PC, while mobile applications might use a laptop or PDA with wireless Internet access through an integrated GPRS modem (see Fig. 1). Efforts to integrate the necessary client-software directly into DGPS/RTK receivers are under development.

Fig. 1: Simultaneous real-time access to DGPS and GIS data by means of Pocket PC (right) and plugged-in GPRS modem (left)

2 Basic Principle

The generation of differential GPS correction data is usually done directly on the GPS receiver of a reference station, but can also be derived from networked observations obtained by a number of reference stations. The data stream thus generated is fed into a server which makes the data accessible on the Internet by means of an appropriate protocol (see Fig. 2). A mobile user might gain access to the Internet via mobile phone technology using a client program which forwards the correction data from

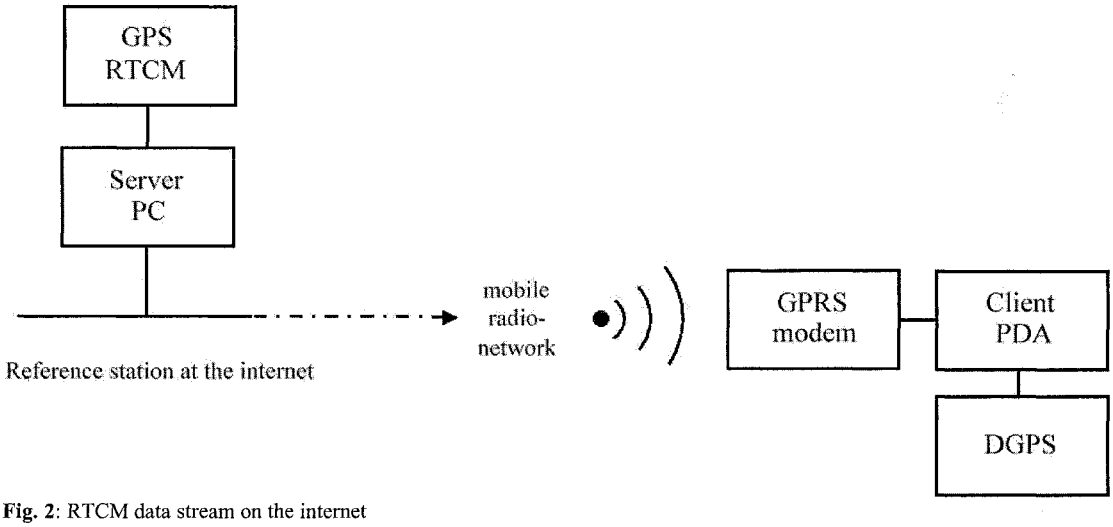


Fig. 2: RTCM data stream on the internet

the server to a GPS receiver. The distance between the reference station and the client with the connected rover station is split into two parts. The larger part of the distance consists of a wired Internet connection, while the remaining part may be bridged using wireless mobile phone technology.

It is up to the users of a DGPS or Real Time Kinematic (RTK) GPS service to choose a technique for retrieving correction data via the Internet. GSM, GPRS, and EDGE are among the choices. The introduction of the „General Packet Radio Service" (GPRS) as an IP packet-switching service for mobile data communication is of particular importance concerning DGPS applications. GPRS strengthens above all the mobile use of the Internet and is generally considered as the precursor of UTMS. In contrast to GSM, the data are divided into packets and transmitted separately, thus allowing optimum use of the bandwidth available within a mobile radio cell. Of decisive importance for DGPS applications is the fact that with GPRS, billing is performed on the basis of the received data volume. This means that the user will no longer pay for the duration of a connection, but for the quantity of data transferred. The available transmission capacity is shared between all participants of a radio cell, while occupying the radio line only when data packets are actually transmitted. Compared to the GSM technique, the costs for the GPRS transmission of DGPS data decreases by up to 90%. However, in the case of RTK applications the use of GPRS unfortunately allows today only a slight cost reduction owing to the tenfold higher data rate.

The Internet (being conceived as an open net) is particularly well suited to transmit data between different providers over long distances. Uniform structures across national borders on a continental or global level can easily be created. However, the servers required must be tied to the Internet via interconnected broadcasters with sufficient bandwidth. Broadcaster operating expenses mainly depend on the number of simultaneously served clients.

Ntrip

The BKG, in cooperation with the University of Dortmund, has developed a http-based (Hyper Text Transfer Protocol) technique for streaming DGPS data, or other kinds of GNSS data, to mobile receivers via the Internet. This technique establishes a format called "Networked Transport of RTCM via Internet Protocol" (Ntrip), because its main application is the dissemination of corrections in the popular RTCM-104 streaming format. Sufficient positioning precision is obtained if correction data are not older than a few seconds. The RTCM standard is used worldwide, and most, if not all, DGPS receivers accept it.

Ntrip's system components are

- NtripSources, which generate DGPS data-streams at a specific location,
- NtripServers, which transfer the data from one or multiple sources in Ntrip format,
- NtripCaster, the major stream-splitting and broadcasting system component,

NtripClients, receiving data of desired sources from the caster.

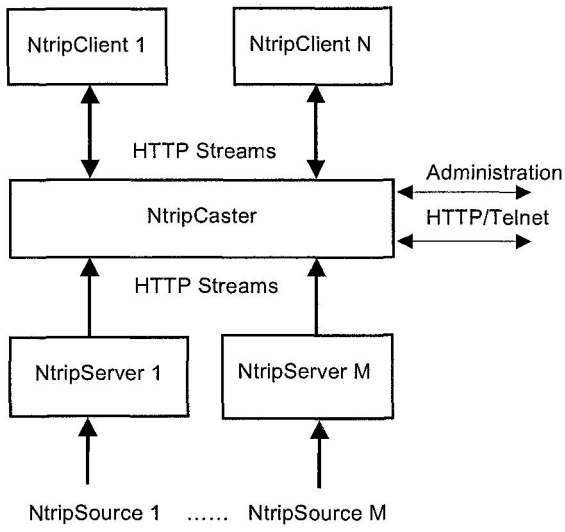


Fig. 3: Ntrip Components

The data streaming is accomplished using the TCP/IP protocol stack. Several attempts, based on Serial-to-TCP conversion of streaming data on the reference-side (server) and TCP-to-Serial re-conversion on the rover-side (client), have demonstrated the suitability of the TCP/IP protocol for streaming data to mobile IP clients. The new technique has two advantages:

a) Ntrip is based on http as an application protocol layer on top of TCP. Consequently any data stream is transmitted exclusively through http's standard port 80. Thus, many problems with firewalls and proxyserver-protected LANs can be avoided.

b) The caster, set up between server and clients as is usual in Internet Radio implementations, duplicates incoming source data, so that they can be received simultaneously by many users. While avoiding NtripServer–NtripClient contacts, the caster furthermore acts as a protecting security element for data stream providers.

Reference stations

As of July 2003, about 90 RTCM correction or raw data streams are available through Ntrip from the reference stations of various institutions. In Germany, a continuously operated network called GREF, comprising about 20 GPS/GLONASS reference stations monitored by BKG on a daily basis, is currently converted to real-time operation. The stations involved will be successively networked over leased lines or Internet connections with a central processing unit. Based on these networked reference stations a service called GREF-IP (IP for Internet Protocol) has been created, which derives DGPS correction data for about 20 virtual reference stations spread over the territory of the Federal Republic of Germany. This service feeds its RTCM data into the Internet through an NtripCaster. Depending on the quality of the GPS receiver on the user side, potential accuracies of up to ± 0.5 m will be available countrywide. The central processing unit uses Trimble's GPSNet-Software for the generation of the correction data.

For establishing DGPS services it is a necessity to offer correction data streams that span national borders. The European Permanent Network (EPN), comprising approx. 160 GPS permanent stations distributed over the whole continent and operated by European National Mapping Agencies (within the scope of the IAG Sub-commission EUREF), constitutes an appropriate datasource for the extension of the GREF-IP concept. In June 2002 EUREF adopted a resolution to provide real-time RTCM data via Internet to serve DGPS positioning and navigation applications. In this way a continental DGPS service is being set up, known as EUREF-IP. Those EPN stations considered for integration into EUREF-IP are operated at close proximity to Internet ports, and use GPS receivers that can generate RCTM correction data. In addition, a number of public or private institutions, have indicated an interest in providing DGPS or RTK data over the Internet. Some of them already make available test data streams. An overview on today's real-time GNSS data availability in Europe is given in Fig. 4.

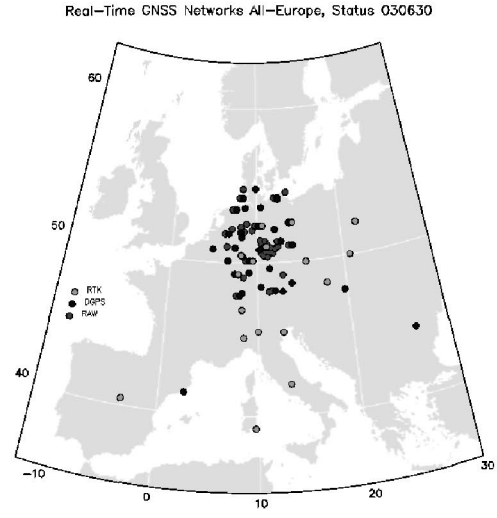


Fig. 4: Real-time data available in Europe through Ntrip, status July 2003

Accuracy and Latency

With any kind of transmission technique the travel time of DGPS or RTK correction data is of decisive importance. The positioning software on the rover side requires the relevant data almost instantaneously (depending, to a certain degree, on the type of positioning). Considerably delayed, missing or irregularly arriving correction data will degrade accuracy, not such a problem when exclusive transmission paths (e.g. AM or FM broadcasters) are used. However, when using the Internet, a DGPS application shares resources of an Internet Service Provider (ISP) with other applications and hence the bandwidth cannot be controlled. The question arises in this context as to what extent the usual resources offered by professional ISPs meet the requirements of real-time positioning and navigation. Fig. 5 presents statistics on the latencies of DGPS and RTK correction data (formats RTCM 2.0 or RTCM 2.1) as derived from tests. The communication over Internet, and wired as well as wireless phone networks (GSM and GPRS), has been investigated. Latencies of the order of less than three to four seconds are typical for the situation in the Federal Republic of Germany. The tests performed in Europe (across national borders) show very few latencies exceeding four seconds. The latency of correction data caused by the inclusion of wireless Internet networks can therefore be considered so low that no noticeable influence on the achievable positioning accuracy exists.

By means of Fig. 6, example-based typical files of 1s-recordings of a static position determination at a known site are illustrated. RTCM corrections were received via the Internet and mobile radio (GPRS). The figure shows differences to a known reference position over a period of one hour. Neither for DGPS nor for RTK applications were the achievable accuracies degraded by RTCM data transmission via mobile communication.

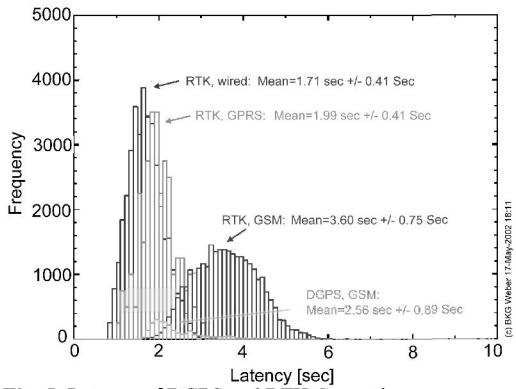


Fig. 5: Latency of DGPS and RTK Corrections

Conclusions

Ntrip is a new technology to transfer GNSS data (for example RTCM correction data) via the Internet and mobile IP networks. Practical tests do not show a significant degradation of performance compared to using other transportation media. Ntrip's main software components are developed within EUREF under the GNU General Public Licence. RTCM's Special Committee 104 has set up a Working Group "Internet Protocol" to further develop Ntrip towards an international standard. Due to its almost worldwide availability and the widespread hardware for wireless Internet access, Ntrip is the streaming method of the future. EUREF operates an NtripCaster. This service has reached Initial Operation Capability. "Notice Advisories to Broadcaster Users" (NABUs) inform users about temporary outages of the service.

References

EUREF-IP Real-Time Project
http://www.epncb.oma.be/projects/euref_IP/euref_IP.html

Distribution Maps of Real-Time GNSS Data

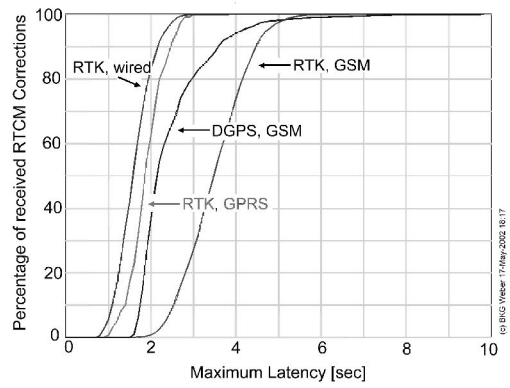
Availability in Europe
http://igs.ifag.de/root_ftp/misc/ntrip/maps

Networked Transport of RTCM via Internet Protocol (Ntrip), Documentation
http://igs.ifag.de/root_ftp/software/NtripDocumentation.zip

Ntrip Software Downloads
http://igs.ifag.de/ntrip_down.htm

EUREF-IP Ntrip Broadcaster
http://igs.ifag.de/index_ntrip_cast.htm

Notice Advisory to Broadcaster Users (NABU)
http://igs.ifag.de/root_ftp/misc/ntrip/nabu



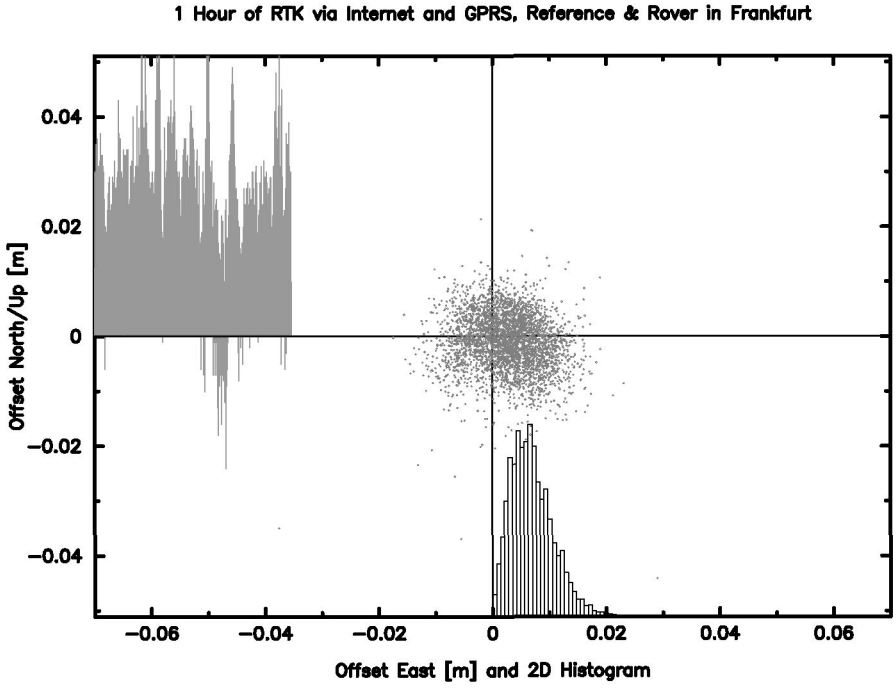


Fig. 6: Accuracy of RTK over Internet and GPRS

Numerical Weather Predictions for GPS Positioning

A.B.O. Jensen
Informatics and Mathematical Modelling, Technical University of Denmark,
Richard Petersens Plads, Building 321, DK-2800 Lyngby, Denmark

Abstract. When GPS satellite signals are transmitted through the atmosphere they are affected by the media. In the neutral atmosphere the refraction is a function of pressure, temperature, and humidity along the signal path, and in the GPS positioning process this effect is normally handled by utilising global tropospheric delay models. For high accuracy differential positioning over baselines lengths where the differential effect of the signal delay from the neutral atmosphere is significant, these global models of the signal delay are not sufficiently accurate, and this is especially the case during abnormal weather conditions.

This paper describes a new approach where numerical weather predictions (NWPs) are introduced in the GPS data processing instead of global tropospheric delay models. NWPs are predictions of the meteorological conditions for a given area and epoch in time, and can as such be used for estimating the tropospheric delay for a satellite signal by numerical integration along the signal path through the NWP. For the tests described in this paper, the signal delays are determined as a zenith delay through the NWP combined with a mapping function. This approach is useful for kinematic and shorter static GPS applications. The paper describes the theory of the method, and the applicability of the method is evaluated by analysing position accuracies obtained by introducing NWP-derived signal delays in kinematic and static processing of GPS data. Improved position accuracies are obtained for most of the test scenarios, indicating that the method does have a potential.

Keywords. GPS positioning, numerical weather predictions (NWP), tropospheric delay

1 Introduction

The GPS satellite signals are affected while being transmitted through the lowest parts of the Earth's atmosphere, the ionosphere and the neutral atmosphere. Normally, global atmospheric models are used to correct for the atmospheric effect, and the models are sufficiently accurate for most GPS positioning with modest accuracy demands. For high accuracy differential carrier phase-based positioning the global models are, however, not sufficiently accurate and other measures must be introduced in the data processing in order to handle the atmospheric effects.

The double differencing technique is commonly used since the technique results in cancellation of satellite and receiver clock errors during the processing. When utilising both atmospheric modelling and double differencing, the residual atmospheric effect is considerably mitigated, and position accuracies of a few cm can be obtained when the distance between the reference and roving GPS receivers is less than approximately 20 km [4].

For longer GPS vectors the ionospheric error can be further mitigated using dual frequency equipment and the so called ionosphere free linear combinations of the L1 and L2 observations [16]. With this technique the first order effects of the ionosphere are removed from the data processing, but also the noise level is increased by generating the linear combinations. For satellite signals received at high elevation angles the size of the residual higher order ionospheric effects and the increased noise are negligible [8], but for signals received at elevation angles below 15 degrees, the effects become significant and should be taken into consideration during the data processing, especially if the GPS vectors are longer than 50-100 km [3].

The effect caused by the neutral atmosphere, normally referred to as the tropospheric delay, is handled by using global tropospheric delay models. When the ionosphere free linear combination is used in the GPS data processing, the residual tropospheric effect after both modelling and double differencing is the dominant error source for carrier phase-based positioning with distances ranging from 20 km to a few hundred km between reference and rover. This is the case under normal conditions when the ionospheric activity is low or moderate. During high ionospheric activity there might be significant higher order ionospheric effects even though the ionosphere free linear combination has been used.

For most geodetic and geodynamic GPS applications the positioning is based on long observation time spans, and the considerable amount of data available leaves enough degrees of freedom in the adjustment process to estimate the residual tropospheric effect as a separate element in the adjustment. This approach is used for instance by the Bernese software [1].

For other applications, such as Real Time Kinematic (RTK) positioning or kinematic post processed positioning there is normally not enough data, from each position solution, to actually estimate the size of the residual tropospheric error. Therefore, other methods for dealing with the residual tropospheric effect must be used. In the following a new approach for estimating the tropospheric delay is described.

2 The tropospheric delay

The tropospheric delay is caused by refraction of the satellite signal in the lowest parts of the Earth's atmosphere, extending from the surface of the Earth up to an altitude of approximately 50 km. This part of the atmosphere, covering the troposphere and the stratosphere, is referred to as the neutral atmosphere by meteorologists since it is electrically neutral to radio waves.

Refraction in the neutral atmosphere causes a signal delay which is normally referred to as the tropospheric delay in GPS terminology, because the majority of the effect occurs in the troposphere. The tropospheric signal delay can be determined by integrating the refractivity (N), along the signal path (ds):

$$\text{delay} = 10^{-6}\int N\,ds \tag{2.1}$$

The refractivity is a function of the meteorological parameters; pressure, temperature and humidity, and can be described by Equation 2.2 as given for instance in [9].

$$N = k_1\left(\frac{P_d}{T}\right)Z_d^{-1} + \left(k_2\frac{e}{T} + k_3\frac{e}{T^3}\right)Z_w^{-1} \tag{2.2}$$

Where k_1, k_2, and k_3 are constants, Pd is the partial pressure of dry air, T is temperature, and e is the partial pressure of water vapor. Z_d and Z_w are the compressibility factors for dry air and water vapor respectively, and they are accounting for the deviation of the gasses from an ideal gas.

By combining Equation 2.1 and 2.2 an expression for the tropospheric signal delay is obtained, and the effect of the signal delay, converted to metric units, is a range error of approximately 2.3 - 2.5 meters in zenith for a GPS receiver located at sea level. The size of the error increases for signals received at lower elevation angles.

Normally the tropospheric delay is handled using global tropospheric delay models as, for instance the Hopfield model [5] or the Saastamoinen model [13]. In 1999 V. Mendes performed a comprehensive analysis of global tropospheric delay models, and he found the Saastamoinen model to be one of the best models, with an estimated accuracy of 3 cm in zenith [10]. For lower elevation angles the model becomes less accurate, indicating the need for either dealing with the residual tropospheric effects in the GPS data processing, or for investigating methods to obtain more accurate estimates of the signal delay. Numerical weather predictions (NWPs) can be used in the second case as described in the following section.

3 NWPs for tropospheric modelling

Numerical Weather Predictions (NWPs) are three dimensional models of the atmospheric conditions for a given area and point in time. The models form the basis for weather predictions and contain the necessary information for estimating the refractivity along the path of a GPS signal.

Equation 2.2 can be rewritten to obtain the following expression:

$$N = k_1 R_d \rho + \frac{R_d}{\varepsilon}\left(k_2 - k_1\varepsilon + \frac{k_3}{T}\right)q\rho \tag{3.1}$$

Where the new variable R_d is the gas constant for dry air, ρ is the density of air, ε is the ratio between the gas constants for dry and wet air, and q is specific humidity. This expression is more suitable for the test purposes described in this paper, since specific humidity is given directly in the weather models used. Further, it is more convenient to perform the integration with respect to air pressure than with respect to altitude, since pressure levels are normally used as the vertical reference in weather predictions. The expression for the signal delay is therefore rewritten, as shown in [7], and Equation 3.2 for the signal delay in the zenith direction is finally obtained.

$$ztd = 10^{-6}\int_{p_{top}}^{p_{ant}} \frac{k_1 R_d}{g} dp + 10^{-6}\int_{p_{top}}^{p_{ant}} \frac{R_d}{\varepsilon}\left(k_2 - \varepsilon k_1 + \frac{k_3}{T}\right)\frac{q}{g} dp \quad (3.2)$$

The integration is carried out from the pressure at the top of the neutral atmosphere, p_{top} down to the pressure level at the GPS antenna, p_{ant}. The new variable in this equation is g which is gravity.

In the derivation of this Equation 3.2 which was first given by H. Vedel in [17], the hydrostatic equation is introduced, thus assuming that the atmosphere is in hydrostatic equilibrium. This is an approximation, which is commonly used for weather modelling, and is expected to be valid for the applications given in this paper.

Based on Equation 3.2 the tropospheric zenith delay can now be derived by extracting pressure, temperature and humidity from a numerical weather prediction. For the following tests the DMI-HIRLAM-E model was used, which is described in [14]. This model is grid based with a 0.15° x 0.15° spacing, 31 vertical layers, and it covers all of Europe. Data from 5. and 14. September 2000 was made available by the Danish Meteorological Institute, and the models were given as one hour predictions.

A routine for estimating tropospheric zenith delay based on Equation 3.2 and the DMI-HIRLAM-E data was developed. The values for temperature and humidity were determined by linear interpolation for each intersection point between the signal path and layers in the NWP. Also the meteorological parameters for the location of the GPS antenna were determined from the nearest layers in the DMI-HIRLAM-E model by a linear interpolation / extrapolation. The contribution to the zenith delay from above the weather model was determined using the expression given in [17]. For the k-constants in Equation 3.2 the values given in [2] have been used.

4 Test of NWP derived zenith delays

The accuracy of the zenith delays derived using the above procedure was verified by a comparison with GPS-derived zenith delays. Twenty four hours of dual frequency GPS data from 14 permanent GPS reference stations located in Denmark and southern Sweden were used to test the accuracy of the NWP derived zenith delays. The GPS data was processed with the Bernese processing software [1], with precise orbits, and accurate station coordinates given in the same reference frame and time epoch as the satellite orbits. The ionospheric effect was handled by introducing the ionosphere free linear combination of the observations, and any errors arising from multipath and receiver noise is expected to be low, because of the high quality equipment installed at the GPS sites. The sites are equipped with choke ring antennas and geodetic grade dual frequency GPS receivers. Residual errors from the GPS data processing are thus only ocean loading effects, which are not modelled in Bernese, but are expected to be negligible for the current stations. Higher order ionospheric effects are also expected to be small, since the ionospheric activity on the observation days was low.

The differences between the GPS derived zenith delays and delays derived from the DMI-HIRLAM-E data have a mean of 0.6 cm, a standard deviation of 1.6 cm and an RMS of 1.7 cm. These results were obtained using data from the 14 sites and 16 instances in time. Similar accuracies of NWP derived zenith delays have been obtained by other groups, as described for instance in [12], [15], and [17]. This indicates that the NWP derived zenith delays are better than the Saastamoinen model, considering its estimated accuracy of 3.0 cm as given in [10].

The use of an improved estimate of the tropospheric delay can improve the accuracy of GPS positions, if the size of the residual errors of all other error sources in the processing is smaller than the size of the improvement in the tropospheric delay estimate. Static and kinematic positioning tests were therefore performed in order to verify whether an improvement in positioning performance could be obtained.

The positioning tests were carried out using commercial GPS processing software, since the method is expected to have a larger impact on shorter static and kinematic processing scenarios which are often handled with commercial software. As mentioned in the introduction, residual tropospheric modelling effects are normally estimated as a part of the positioning process for long static observation campaigns, where the processing is carried out using more advanced software.

5 Positioning tests

The positioning tests were carried out using the GPS data from the 14 permanent GPS stations mentioned above and shown in Figure 1. Two six hours data periods from each day were processed with a 15 second data rate. Precise orbits were introduced, and the ionosphere free linear combination of the observations was used to

eliminate the majority of the ionospheric effect. The data was processed using the GPSurvey software version 2.35 from Trimble Navigation.

Station ONSA was used as reference station, and the resulting GPS vector lengths ranged from 35 to 288 km. The largest height difference between reference and roving receiver was 215 meters.

Before processing the data the new tropospheric correction approach was applied to the RINEX GPS data files. Initially zenith delays were estimated from the DMI-HIRLAM-E data for each GPS position at every full hour where weather data was available. A linear temporal interpolation was then carried out in order to estimate the zenith delay for every data epoch. The Niell mapping function [11] was applied to the zenith delays, to determine the slant tropospheric delay, and finally the slant delay was subtracted from the raw code and phase observations in the GPS data files. Tropospherically corrected RINEX GPS data was hereby generated.

The processing was carried out twice firstly with the tropospherically corrected GPS data and then with the Saastamoinen tropospheric correction enabled in the software. All other parameters were identical. The processing was carried out in both static and kinematic mode.

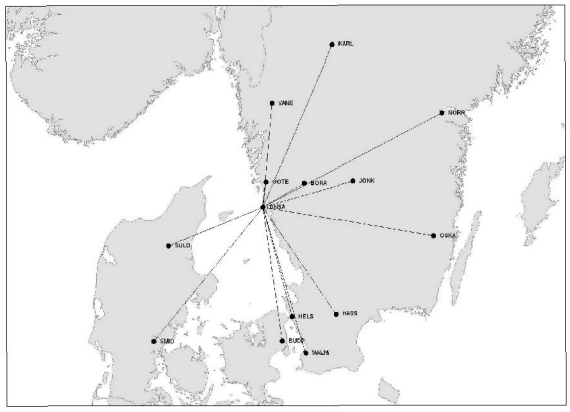


Fig. 1. Location of GPS stations, and vectors used for the positioning tests

5.1 Static positioning results

The resulting coordinates from the static processing were compared to the previously known coordinates for the GPS stations. In Table 1 mean and standard deviations are given for the 3D coordinate differences with respect to the previously known stations coordinates, i.e. the size of the vector between the position given by the processing result and the known position. All mean values are therefore positive in the table. It was expected that the effect of using the NWP derived zenith delays would be greater with a larger separation between the reference and roving GPS receivers, or with a larger height difference between the two receivers. This was, however, not the case, and the results of the different static baseline processings are therefore grouped together in the table.

Table 1. Statistics from static positioning test

Static results	Mean and standard deviation of 3D position difference	
	Saastamoinen	NWP approach
5. Sept. 2000		
00:00–06:00	0.135 ± 0.113 m	0.115 ± 0.068 m
12:00-18:00	0.092 ± 0.076 m	0.086 ± 0.059 m
14. Sept. 2000		
00:00-06:00	0.037 ± 0.032 m	0.020 ± 0.014 m
12:00-18:00	0.052 ± 0.044 m	0.112 ± 0.070 m

Table 1 shows that both mean and standard deviation are generally smaller when using the NWP approach. The position accuracy was, however, only improved with the NWP approach for 25 of the 51 positions determined, and for the afternoon data set of 14. September the Saastamoinen model outperforms the NWP approach.

During 14. September a low pressure system was moving over the area creating a relatively high tropospheric activity, and it is expected that this is causing problems with the tropospherically corrected data files during the data preprocessing in GPSurvey. Considering the speed of the GPS satellites and the Niell mapping function, a change in tropospheric slant delay of approximately 0.5 meter occurs when a satellites rises with from a 5.0° to a 5.1° elevation angle, if the zenith delay is 2.4 meters [7]. Such a change in elevation angle can take place during 15 seconds, i.e. in the present case from one data epoch to the next. With the tropospherically corrected data files a change in the phase observations of more than two cycles will be interpreted as a cycle slip when the data is imported by GPSurvey. Therefore a 15° elevation mask was used for the data processing, whereby most of this effect was eliminated for the results given in Table 1. There might, however, have been some remnants of this effect left in the data set collected during the afternoon of 14. September, and it is expected that filtering carried out by GPSurvey, during pre-processing of the data, is accidentally removing parts of the tropospheric corrections. Hereby worse results for the NWP approach are obtained with the afternoon data of 14. September.

The ionospheric activity on the two days was low. But in the afternoon and evening hours of 4. September, the ionospheric activity was high, and this may have affected the GPS data collected in the morning hours of 5. September, resulting in the relatively large mean values given in the second row in Table 1.

5.2 Kinematic positioning results

For the kinematic test, the static data sets from the 14 GPS stations were processed in kinematic mode, so one position was determined for each station every 15 second. "On-the-fly" ambiguity resolution was used with the optimal search algorithm implemented in the GPSurvey software. All other processing parameters were the same as for the static processing mentioned above.

The 1440 positions obtained from positioning each of the 13 "roving" GPS receivers through the six hour periods were subtracted from the previously known coordinates of the GPS stations, and the mean and standard deviation of the size of the 3D error vector, determined as for the static case, are shown in Table 2. Positions determined for time epochs where the RDOP was higher than 4 are not included in the results given in Table 2.

The results of the data from 5. September show an improvement in both mean and standard deviation when using the NWP approach. For the morning of 14. September the results are almost identical for the two tropospheric approaches, but for the afternoon hours the best results are obtained with the Saamstamoinen model. The explanation is expected to be the same as for the static case, i.e. difficulties in GPSurvey with handling large variations in tropospheric delay during data import.

Table 2. Statistics from kinematic positioning test

Kinematic results	Mean and standard deviation of 3D position difference	
	Saastamoinen	NWP approach
5. Sept. 2000		
00:00–06:00	0.239 ± 0.223 m	0.231 ± 0.169 m
12:00-18:00	0.200 ± 0.122 m	0.156 ± 0.096 m
14. Sept. 2000		
00:00-06:00	0.098 ± 0.103 m	0.101 ± 0.102 m
12:00-18:00	0.072 ± 0.088 m	0.144 ± 0.092 m

Given in Table 3 is the number of positions determined for each of the kinematic processing scenarios.

As shown in Table 3 there is a relatively large difference in the number of positions determined with the two tropospheric approaches. The GPSurvey software does not determine a kinematic position for a given time epoch if the residual noise level in the processing is too high, and the difference in number of positions determined is therefore expected to be caused by differences in the residual noise level. Generally more positions were determined with the NWP approach, and this could be the result of a better modelling of the tropospheric delay, where the amount of unmodelled tropospheric effects is decreased implying a lower noise level in the data processing.

Table 3. Positions determined from kinematic test

Kinematic results	Total number of positions determined	
	Saastamoinen	NWP approach
5. Sept. 2000		
00:00–06:00	13559	17116
12:00-18:00	16007	17452
14. Sept. 2000		
00:00-06:00	17298	18181
12:00-18:00	13768	14286

It is also interesting to investigate how the ambiguity resolution is affected by the different approaches for tropospheric corrections. Of the positions used for Table 2 and 3 the percentage of positions based on fixed ambiguities are shown in Table 4.

Table 4. Percentage fixed solutions

Kinematic results	Percentage of positions based on fixed ambiguities	
	Saastamoinen	NWP approach
5. Sept. 2000		
00:00–06:00	84 %	98 %
12:00-18:00	99 %	100 %
14. Sept. 2000		
00:00-06:00	99 %	100 %
12:00-18:00	100 %	100 %

Generally the number of fixed ambiguities are correlated with the position accuracies as given in Table 2, where the statistics is based on both fixed and float solutions. As expected, with the relatively long baseline lengths processed in this case, some ambiguities were fixed to wrong integer values. Previous experience with the software has shown that it does have problems with solving ambiguities to the correct integer values, when processing kinematic data from longer baselines [6].

6 Conclusion and future work

A new approach for estimating the tropospheric delay based on numerical weather predictions has been described in this paper. Tests indicate that the accuracy of tropospheric zenith delays based on the

DMI-HIRLAM-E data are better than the accuracy of the Saastamoinen global tropospheric delay model. The static and kinematic positioning tests also show that the method does have a potential for improving GPS positioning performance, and especially with the first two data sets there is a general improvement in position accuracy for both the static and kinematic results.

However, for the data collected during the period with high tropospheric activity, the positioning performance is generally not improved with the NWP approach. This is in contrast to what was expected, since a better modelling of the tropospheric delay was expected to be more beneficial during times with larger variability in the delay values. The explanation is assumed to be found in the processing software used for the tests. Using another test setup, where the tropospheric correction routine can be built directly into the source code of the programme, is therefore necessary in order to solve this problem. Also tests with more data, for instance from other climate regions, or from days with high ionospheric activity, are necessary to finally conclude whether the method is feasible or not.

The new technique is dependent on the accuracy of the weather models and tests should also be carried out using other weather models, for instance with a smaller or larger grid spacing, in order to investigate how this will affect the positioning results.

7 Acknowledgements

The Danish Meteorological Institute (DMI) is acknowledged for the DMI-HIRLAM-E data, and Henrik Vedel from DMI is thanked for advice and discussions related to the use of the data. The National Land Survey of Sweden is acknowledged for the SWEPOS GPS data, and the National Survey and Cadastre - Denmark (KMS) is acknowledged for the Danish GPS data. Bo Madsen from KMS is thanked for helping with the Bernese data processing.

References

1. Beutler, G., E. Brockmann, R. Dach, P. Fridez, W. Gurtner, U. Hugentobler, J. Johnson, L. Mervart, M. Rothacher, S. Schaer, T. Springer, R. Weber (2000). Bernese GPS Software. Astronomical Institute, University of Berne.
2. Bevis, M., S. Businger, S. Chriswell, T. A. Herring, R. A. Anthes, C. Rocken, R. H. Ware (1994). GPS Meteorology: Mapping Zenith Wet Delays onto Precipitable Water. Journal of Applied Meteorology, 33: 379-386.
3. Brunner, F. K., M. Gu (1991). An improved model for the dual frequency ionospheric correction of GPS observations. Manuscripta Geodaetica, 16:205-214.
4. Cannon, M. E. (1997). Carrier Phase Kinematic Positioning: Fundamentals and Applications. In Geodetic Applications of GPS, Lecture Notes for Nordic Autumn School edited by Bo Jonsson. Number 16 in Reports in Geodesy and Geographical Information Systems, National Land Survey of Sweden.
5. Hopfield, H.S.(1969). Two-quartic Tropospheric Refractivity Profile for Correcting Satellite Data. Journal of Geophysical Research, 74(18): 4487-4499.
6. Jensen, A. B. O., M. E. Cannon (2000). Performance of Network RTK Using Fixed and Float Ambiguities. Proceedings of the 2000 National Technical Meeting of the Satellite Division of the Institute of Navigation (ION NTM 2000). Pages 797-805.
7. Jensen, A. B. O. (2002). Numerical Weather Predictions for Network RTK. Publication Series 4, volume 10. National Survey and Cadastre – Denmark.
8. Johansson, J. M. (1997), Modelling of the Earth Atmosphere in Space Geodetic Applications. In Geodetic Applications of GPS, Lecture Notes for Nordic Autumn School edited by Bo Jonsson. Number 16 in Reports in Geodesy and Geographical Information Systems, National Land Survey of Sweden.
9. Langley, R. (1996). Propagation of the GPS Signal. In Kleusberg, A. and P. J. G.Teunissen (eds) GPS for Geodesy, Lecture Notes in Earth Sciences. Springer-Verlag.
10. Mendes, V. B. (1999). Modelling the neutral-atmosphere propagation delay in radiometric space techniques. Ph.D. dissertation. Report number 199. Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredricton.
11. Niell, A. E. (1996). Global mapping functions for the atmosphere delay at radio wavelengths. Journal of Geophysical Research, 101(B2):3227-3246.
12. Pany, T., P. Pesec, G. Stangl, (2001). Atmospheric GPS Slant Path Delays and Ray Tracing Through Numerical Weather Models, a Comparison. Physics and Chemistry of the Earth. 26A(3):183-188.
13. Saastamoinen, J. (1973). Contributions to the Theory of Atmospheric Refraction. Bulletin Geodesique. Printed in three parts, 105:279-298, 106:383-397, 107:13-34.
14. Sass, B. H., N. W. Nielsen, J. U. Jørgensen, B. Amstrup, M. Kmit (2000). The Operational HIRLAM System at DMI. Scientific Report 00-26. Danish Meteorological Institute. Copenhagen
15. Schueler, T. (2001). On Ground-based GPS Tropospheric Delay Estimation. Ph.D. thesis, Universität der Bundeswehr, München.
16. Seeber, G. (1993). Satellite Geodesy. Foundations, Methods and Applications. Walter de Gruyter.
17. Vedel, H., K. S. Mogensen, X.-Y. Huang (2001). Calculation of zenith delays from meteorological data, comparison of NWP model, radiosonde and GPS delays. Physics and Chemistry of the Earth, 26A(6-8):497-502.

A Multi-Base-Station Approach for Long Range Differential GNSS Positioning of Airborne Sensors

N. S. Kjorsvik, O. Ovstedal and J. G. G. Svendsen
Dept. of Mapping Sciences, P.O.Box 5034, Agricultural University of Norway, N-1432 As, Norway

L. E. Blankenberg
Norkart as, P.O. Box 145, N-1300 Sandvika, Norway

Abstract. In airborne applications (e.g. photogrammetry, gravimetric measurements and laser scanning) the positions of the sensors are determined by Global Navigation Satellite Systems (GNSS). Traditionally, this has been done with differential techniques using GNSS observations from one reference station only. The distance and height dependent errors may be significant in airborne differential positioning using one reference station only. In this work the alternative approach of Multi-Base-Station (MBS) differential processing is discussed and verified. In the MBS approach, data from a network of permanent reference stations are used to estimate the distance dependent errors. Improved reference data can be materialized as e.g. observations from a Virtual Reference Station (VRS). Results from a test flight over a dedicated test field in Fredrikstad, Norway, are presented. Reference positions are estimated using photogrammetric aerial triangulation and compared to the positions from various GNSS processing strategies. The results verify that the method of Multi-Base-Station processing gives significantly more precise and reliable results than the traditional approach using one base station only.

Keywords. GNSS, Virtual Reference Stations, Airborne Sensors

1 Introduction

Airborne sensors are commonly used in mapping applications. In, e.g., GNSS-supported aerial triangulation, a significant number of Ground Control Points (GCP) can be replaced by GNSS-determined antenna positions. A special advantage in GNSS-supported aerial triangulation is the possibility to estimate systematic errors in the GNSS-derived antenna coordinates. When having a minimum of four GCPs in a rectangular photogrammetric block, the accuracy of estimated coordinates is at the sub-decimeter level when operating up to approximately 500km from the GNSS reference stations (see e.g. Blankenberg, 2002). In other airborne applications, like laser scanning or airborne gravimetric missions, the sensor positions are based on satellite positioning alone, or in combination with inertial systems.

For airborne GNSS receivers operated beyond a few tens of kilometers from a single reference station, the differential errors can make the determination of correct integer carrier phase ambiguities a difficult and unreliable task, thus preventing the full accuracy potential of kinematic positioning with fixed ambiguities. The alternative approach of estimating the coordinates with float carrier phase ambiguities has the drawback of reduced geometric strength due to the additional parameters. The effect of the weaker geometry and systematic differential errors prevents sub-decimeter accuracy. The geometry of a float solution can be improved by increasing the time span of continuous GNSS observations, but in many applications this is not desirable.

Regional GNSS reference networks allows the alternative approach of modeling the errors over a region, followed by correction parameters made available to the user (see e.g. Wanninger, 2000; Raquet, 1998). The aim is centimeter level positioning using carrier phase observables.

In the present paper the concept of Multi-Base-Station differential GNSS is outlined. In the first step the observations from the permanent reference stations are used to estimate distance dependent errors. The distance dependent errors are interpolated into the position of the unknown receiver and applied as corrections in the form of, e.g., a Virtual Reference Station (Wanninger, 1997). The integer double-difference ambiguities are resolved for baselines in the permanent GNSS network. In order to address the distance dependent errors to specific receivers and satellites, an approach to decompose the double-differences to "Quasi Zero-Differences" is outlined. The suggested method has the advantage of being consistent for arbitrary choices of reference stations and reference satellites. Results from a pho-

togrammetric test flight over a dedicated test field in Fredrikstad, Norway, are presented. A reference trajectory was estimated using a local reference station in the test field and verified by comparing with independently determined camera positions from a traditional photogrammetric aerial triangulation using ground control points.

2 Estimation of the distance dependent errors

The GPS double-differenced (DD) carrier phase observation Φ may be expressed in units of length as:

$$\nabla\Delta\Phi_{i,j}^{n,r} = \nabla\Delta\rho_{i,j}^{n,r} + \lambda\nabla\Delta N_{i,j}^{n,r} - \nabla\Delta I_{i,j}^{n,r} + \nabla\Delta T_{i,j}^{n,r} + \nabla\Delta E_{i,j}^{n,r} + \nabla\Delta M_{i,j}^{n,r} + \nabla\Delta v_{i,j}^{n,r}, \quad (1)$$

where

i, j	are receiver indexes
n, r	are satellite indexes
$\nabla\Delta\rho_{i,j}^{n,r}$	is DD geometric range (m)
λ	is wavelength (m/cycle)
$\nabla\Delta N_{i,j}^{n,r}$	is initial DD ambiguity (cycles)
$\nabla\Delta I_{i,j}^{n,r}$	is DD ionospheric delay (m)
$\nabla\Delta T_{i,j}^{n,r}$	is DD tropospheric delay (m) after removal of a priori model
$\nabla\Delta E_{i,j}^{n,r}$	is DD ephemeris error (m)
$\nabla\Delta M_{i,j}^{n,r}$	is DD multipath (m)
$\nabla\Delta v_{i,j}^{n,r}$	is DD noise and unmodeled effects (m)

When the integer double-differenced ambiguities ($\nabla\Delta N_{i,j}^{n,r}$) have been determined, the double-differenced carrier phase observable may be regarded as precise range observations.

By omitting satellite and receiver indexes, an "observation minus range" quantity may be defined as

$$\overline{\nabla\Delta\Phi} \equiv \nabla\Delta\Phi - \nabla\Delta\rho - \lambda\nabla\Delta N. \quad (2)$$

$\overline{\nabla\Delta\Phi}$ will now contain the the sum of all errors and noise, as observed on one of the carrier frequencies.

2.1 From Double-Differences to Undifferenced Quantities

While $\overline{\nabla\Delta\Phi}$ is a double-differenced quantity, the goal is to provide one estimate per station and satellite at each observation epoch. Several approaches have been suggested (e.g. Wanninger, 2000; Alber et al., 2000; Raquet, 1998) all of which preserve consistency on a double-differenced level. The following approach is based on Kjorsvik (2003).

The double-differenced terms are transformed to a "Quasi Single-Differenced" (QSD) level by adding an arbitrary constant to each baseline according to eq. (3). The QSDs are computed at a (between stations) single-difference level, but are only consistent on a DD level, hence the term "Quasi" SD:

$$\overline{\Delta\psi_{i,j}^{n}} \equiv \begin{cases} \delta_{i,j}, & n = r \\ \overline{\nabla\Delta\Phi_{i,j}^{n,r}} + \delta_{i,j}, & \text{otherwise,} \end{cases} \quad (3)$$

where $\delta_{i,j}$ is a baseline specific parameter and r is the reference satellite.

The QSDs are transformed to a zero-differenced level by choosing one arbitrary station as master reference and assigning arbitrary constant values κ^n to all its QSDs. "Baseline chains" are then used to form linear combinations of the QSDs to refer them to this master reference station. The resulting "Quasi Zero-Differences" (QZD) $\overline{\psi_i^n}$ of the observation minus range quantities are then stored for each epoch and station in the network as:

$$\overline{\psi_i^n} \equiv \begin{cases} \kappa^n, & i = j \\ \overline{\Delta\psi_{i,j}^{n}} + \kappa^n, & \text{otherwise,} \end{cases} \quad (4)$$

where j denotes one arbitrarily chosen master reference station and κ^n is a satellite specific parameter.

The dual-frequency QZDs may further be separated into a dispersive and a non-dispersive part using the well known relation (satellite and receiver indexes omitted):

$$\overline{\psi_I} = -\frac{\overline{\psi_1} - \overline{\psi_2}}{1 - \frac{f_1^2}{f_2^2}}, \quad (5)$$

$$\overline{\psi_T} = \overline{\psi_1} + \overline{\psi_I}, \quad (6)$$

where subscripts 1 and 2 and the terms f_1 and f_2 indicate $L1$ and $L2$ frequencies, and subscripts I and T indicate the dispersive and non-dispersive parts respectively.

3 Implementation

3.1 Network Ambiguity Resolution

Based on an algorithm in Mervart (1995) an optimal set of independent baselines between the reference stations is automatically identified. The processing is carried out separately for each baseline using double-differenced observations. The first step is the ambiguity resolution for the baselines between the reference stations. The current implementation features

a classical widelane/narrowlane approach (see e.g. Mervart, 1995) and a sequential least squares search.

A priori tropospheric delay is computed using a modified Hopfield model (see e.g. Leick, 1995) and the Niell mapping functions (Niell, 1996). Throughout all processing, relative antenna calibration parameters made available by the US National Geodetic Survey (Mader, 2003) are used to minimize antenna specific effects. In the current implementation, all parameter estimation is carried out in batch least squares adjustments.

3.2 The Virtual Reference Station

A main issue in the construction of a VRS is how to deal with the satellite-specific biases. The two strategies are i) estimation and ii) the use of inherent biases. The first approach (e.g. Jäggi et al., 2001; Wübbena et al., 2001) will hardly be successful for small networks and limited time spans of data. The second approach is to use the observations from one of the reference receivers as these observations naturally contain the satellite specific biases.

The latter method is here preferred for the sake of it's simplicity. The Virtual Reference Station will then inherit the multipath and noise characteristics from the physical observations of the selected master reference station.

Using the carrier phase derived estimates of the distance dependent errors, it is possible to average pseudorange observations between stations. Using n reference stations and considering the combined pseudorange multipath and noise as a stochastic parameter, a $\frac{1}{\sqrt{n}}$ reduction in standard deviation of the averaged pseudoranges can be expected. However, the multipath at each station will not have a zero expectation, yielding slightly worse results. During this process a highly effective outlier detection is possible.

A similar approach may not be used for the carrier phase observables, since the estimates of the distance dependent errors already contain the carrier phase multipath and noise. A reduction in multipath and noise of the carrier phase observables can be accomplished by making assumptions about the behavior of the distance dependent errors. Two cases may be distinguished, i) estimation of parameters assuming functional and/or stochastic models (e.g. Wübbena et al., 2001); ii) short baseline leading to zero differential errors (Wanninger, 2002).

In our implementation the generation of a Virtual Reference Station consists of the following steps:

- Estimation of differential distance dependent errors for each baseline and station based on carrier phase observables.
- Generation of a master reference station with improved pseudorange observations.
- Geometric re-localization and correction for antenna phase center variations.
- Interpolation and correction for the spatially correlated errors. In the current implementation a 2D linear interpolation is applied.

4 Test Flight

A flight was carried out on May 8th 2002 in order to test the proposed methods on real data. Figures 1 and 2 show the trajectory and height profile of this flight.

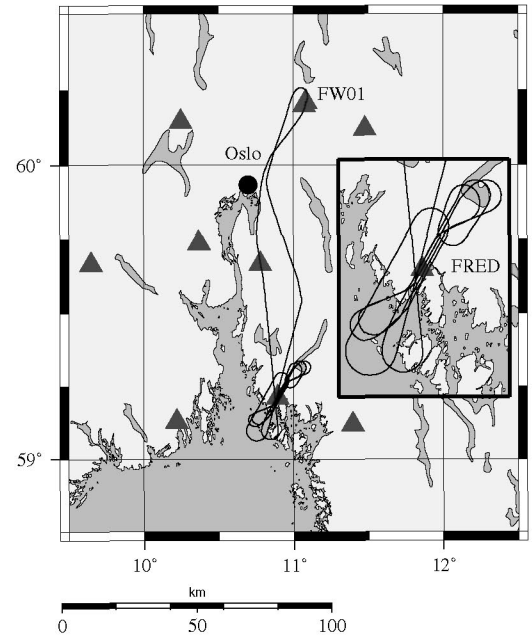


Fig. 1. Trajectory of test flight carried out on May 8th 2002. Reference stations are marked with triangles.

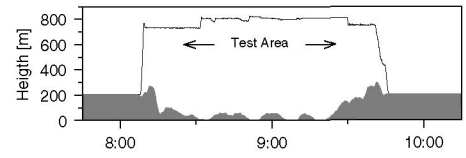


Fig. 2. Height and terrain profile of the test flight.

Dual-frequency observations were recorded at 2Hz in the aircraft and some of the reference stations, and at 1Hz at the remaining reference stations. The reference stations have coordinates in the ETRF89 refer-

ence frame with an accuracy in terms of standard deviations of approximately 0.01m. The company carrying out the test flight operates the reference station FW01 at the airport. Observations from FW01 are routinely used to post-process kinematic GNSS for distances up to 500km in GNSS-supported aerial triangulation (Ovstedal, 2001). The photogrammetric camera in the aircraft is fitted with a wide angle lens cone (c=0.15m), and the altitude over the test field is approximately 800m. The instrumentation and processing of photogrammetrically-derived positions are according to Blankenberg (2002). The estimated precision (standard deviations) of the positions derived by photogrammetry are 0.08m for the North and East coordinates and 0.06m for height coordinates.

The reference station FRED is situated in the midle of the photogrammetric test field, and is used to estimate the reference ("true") trajectory. The baseline FW01 – FRED is approximately 112km. In order to quantify the magnitude of the ionospheric delay, carrier phase observations with fixed ambiguities are used in eq. (5) to estimate ionospheric delays for double-differences. Figure 3 shows the estimated ionosheric delay for the baseline FW01 – FRED. Corrections for the distance dependent errors were interpolated to the station FRED and compared with results from the network solution with fixed ambiguities. Figure 4 shows that the distance dependent errors can be linearly interpolated to a few centimeters for a station on the ground and this reference station spacing.

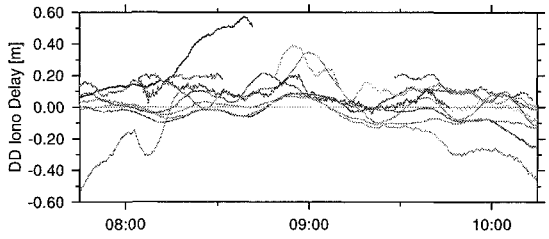


Fig. 3. Differential ionosphere for the baseline FW01 – FRED (112km). Elevation cutoff angle is 10°.

4.1 Reference Trajectory

A kinematic processing with on-the-fly (OTF) ambiguity resolution was carried out with the in-house developed software GPSPROG (Ovstedal, 2001) using the FRED reference station and ionospheric-free linear combinations of L1 and L2 observations. In this processing the time span from UT 08:25 to UT 09:25 was selected where the aircraft was flying over

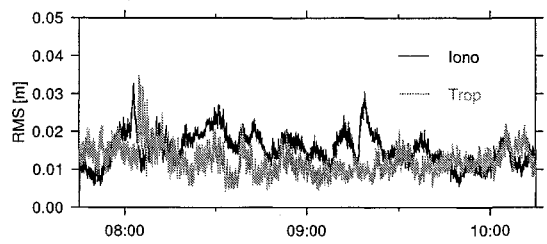


Fig. 4. RMS of the interpolation error at each observation epoch for station FRED. Elevation cutoff angle is 10°.

the testfield in a distance of less than 17km from FRED. This reference trajectory was verified by processing the data using POSGPS 4.02 from Applanix Corporation (Canada) and SKI Pro 3.0 from Leica Geosystems (Switzerland). The differences between the three set of coordinates are less than 0.01m for North and East coordinates. Random differences up to 0.08m can be seen for the height coordinates. One possible reason for the differences might be the use or not use of a dynamic model in the different softwares. Significant air turbulence was experienced during the flight.

The estimated positions from GPSPROG are compared with independently determined positions from the aerial triangulation. In table 1 some statistics are given showing agreement at the sub-decimeter level.

Table 1. Statistics using the local reference receiver (FRED) compared to the photogrammetric solution at the 48 camera exposures.

	North (m)	East (m)	Height (m)
Maximum	0.189	0.265	0.177
Mean	0.003	-0.006	-0.067
RMS	0.073	0.086	0.085

We suspect a systematic height bias due to residual troposphere in the GPS solution. The reference trajectory will thus not be significantly more accurate in height than the cases we will discuss in the following sections.

4.2 Evaluation of the MBS approach

The kinematic data are processed using two different reference stations and compared to the reference trajectory. First the reference station FW01 is used, which would be the normal case for the company carrying out the photogrammetric flight. In the second processing a Virtual Reference Station was generated using all network stations (except FRED). The full time span of available kinematic data from UT 07:58 to UT 09:57 was processed.

In order to quantify the residual effects of both dispersive and non-dispersive errors, single-frequency observations were used in the processing. In the float solutions both kinematic coordinates and the carrier phase ambiguities are estimated in batch least squares adjustments, where coordinates are pre-eliminated in a first program run and estimated in a second program run (Ovstedal, 2001). In the estimation of ambiguities the carrier phase and pseudorange observations are given apriori standard deviations of 0.01m and 1.00m respectively. In the fixed solutions the L1 carrier phase ambiguities are constrained to their correct integer numbers. The correct integer ambiguities are estimated with a static initialization at the airport and an OTF ambiguity resolution for FW01 and the Virtual Reference station, respectively.

The reduction in variance of unit weight of the float solutions is considered an indicator of reduced unmodeled effects.

4.3 Results

Tables 2 and 3 give the statistics for the trajectory estimated with local reference station FW01 for the float and fix solution, respectively. The standard deviation of unit weight in the float solution was 5.26. Figures 5 and 6 presents time series for differences from the reference trajectory. As to be expected, constraining the integer ambiguities improves the accuracy, but systematic errors in the decimeter range can still be seen.

Tables 4 and 5 give the corresponding statistics for the solutions using the Virtual Reference Station. The standard deviation of unit weight for the float solution was 0.97.

Table 6 gives the differences between estimated float ambiguities and correct integer ambiguities for the two solutions. In the float solution using the Virtual Reference Station, the differences are at the 0.1 cycle level. The time series of errors for the float and fix solution are thus almost identical, and figure 7 presents the time series for the float solution.

The accuracy of the fix solution using single-frequency observations and the Virtual Reference Station is at the few centimeter level for horizontal coordinates. It appears to be a bias in the height coordinates which probably can be explained by residual tropospheric effects between the aircraft and the reference receivers on the ground.

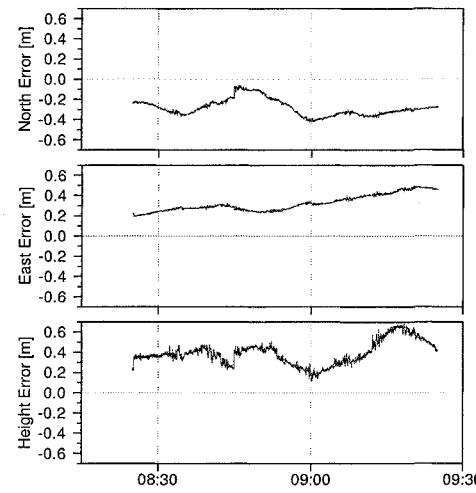


Fig. 5. Position discrepancies for the L1 float solution using the remote reference receiver FW01.

Table 2. Statistics of the L1 float solution using the remote reference receiver FW01.

	North (m)	East (m)	Height (m)
Maximum	0.4250	0.4950	0.6820
Mean	-0.2818	0.3236	0.3917
RMS	0.2933	0.3336	0.4087

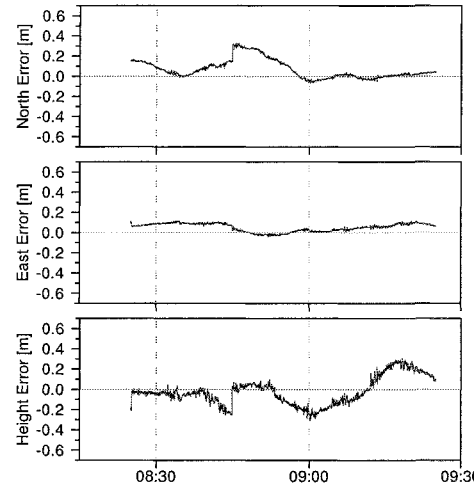


Fig. 6. Position discrepancies for the L1 fix solution using the remote reference receiver FW01.

Table 3. Statistics of the L1 fix solution using the remote reference receiver FW01.

	North (m)	East (m)	Height (m)
Maximum	0.3260	0.1170	0.3080
Mean	0.0699	0.0518	-0.0226
RMS	0.1170	0.0657	0.1386

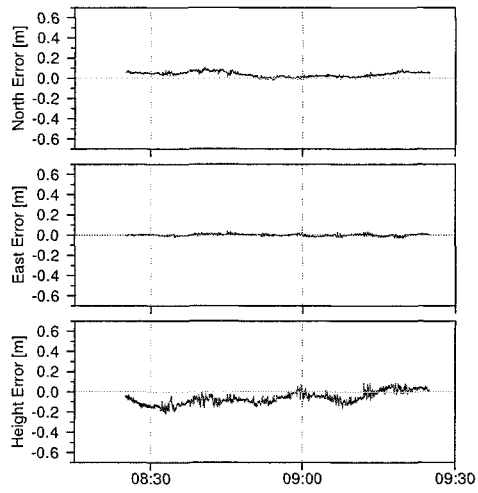


Fig. 7. Position discrepancies for the L1 float solution using the VRS.

Table 4. Statistics of the L1 float solution using the VRS.

	North (m)	East (m)	Height (m)
Maximum	0.1040	0.0360	0.2220
Mean	0.0390	0.0003	-0.0618
RMS	0.0456	0.0092	0.0858

Table 5. Statistics of the L1 fix solution using the VRS.

	North (m)	East (m)	Height (m)
Maximum	0.0830	0.0350	0.2210
Mean	0.0207	-0.0024	-0.0510
RMS	0.0309	0.0098	0.0816

Table 6. Differences of the estimated float ambiguities for PRNs 02 – 26 compared to their correct integer values using FW01 (FW) and the Virtual Reference Station (VR). PRN15 is reference satellite. All differences are given in units of cycles.

	02	03	06	10	17	18	21	22	23	26
FW	1.4	1.5	1.3	3.5	0.2	0.2	0.7	0.3	1.5	1.5
VR	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.1	0.1

5 Conclusions

The concept of Multi-Base-Station differential positioning have been outlined, with special attention to implementation issues. An operational software to generate GPS observation files for Virtual Reference Stations and kinematic rovers has been implemented, and the software is verified using observations from a photogrammetric test flight. Based on the current approach, a significant gain in accuracy for single-frequency solutions was obtained. The increased accuracy for single-frequency solutions will improve both ambiguity resolution and accuracy of the final kinematic solution.

Future research goals include evaluation of more sophisticated modeling of spatial and temporal errors. For airborne applications, the modeling of vertical trophosphere gradients is necessary in order to increase the accuracy of height coordinates.

Acknowledgements. The authors would like to thank the Norwegian Mapping Authority for making parts of the reference station data available and the company FW Geo AS for the airborne observations.

References

Alber, C., Ware, R., Rocken, C., & Braun, J. (2000). Obtaining single path phase delays from GPS double differences. *Geophysical Research Letters*, **27**(17), 2661–2664.

Blankenberg, L. (2002). *GPS-supported aerial triangulation in theory and practice*. Ph.D. thesis, Agricultural University of Norway. ISBN 8257504556.

Jäggi, A., Beutler, G., & Hugentobler, U. (2001). Using double difference information from network solutions to generate observations for a virtual GPS reference receiver. Presented at IAG 2001 Scietific Assembly. Budapest, Hungary.

Kjorsvik, N. (2003). Analysis of data from permanent GNSS networks. Ph.D. thesis (in preparation).

Leick, A. (1995). *GPS Satellite Surveying*. John Wiley and Sons, Inc., 2nd edition. ISBN 0471306266.

Mader, G. (2003). GPS Antenna Calibration at the National Geodetic Survey. http://www.ngs.noaa.gov/ANTCAL/Files/summary.html. Last visited June 16th 2003.

Mervart, L. (1995). *Ambiguity Resolution Techniques in Geodetic and Geodynamic Applications of the Global Positioning System*. Ph.D. thesis, University of Berne, Switzerland.

Niell, A. E. (1996). Global mapping functions for the atmosphere delay at radio wavelengths. *Journal of Geophysical Research*, **101**(B2), 3227–3246.

Ovstedal, O. (2001). *A study of some aspects concerning georeferencing with the Global Positioning System*. Ph.D. thesis, Agricultural University of Norway. ISBN 8257504505.

Raquet, J. (1998). *Development of a method for kinematic GPS carrier phase ambiguity resolution using multiple reference receivers*. Ph.D. thesis, University of Calgary.

Wanninger, L. (1997). Real-Time Differential Error Modelling in Regional Reference Station Networks. In *IAG Symposia Vol 118: Advances in Positioning and Reference Frames*, pages 86–92.

Wanninger, L. (2000). *Präzise Positionierung in regionalen GPS-Referenzstationsnetzen*. Number 508 in Reihe C. Deutschen Geodätischen Kommission, München.

Wanninger, L. (2002). Aus zwei macht eins: Kombination der Beobachtungen von GPS-Referenzstationspaaren. *Allgemeine Vermessungs-Nachrichten*, **109**(10), 352–358.

Wübbena, G., Bagge, A., & Schmitz, M. (2001). RTK Networks based on Geo++ GNSMART (R), Concepts, Implementation, Results. In *Proceedings of ION GPS 2001*, Salt Lake City, Utah, USA. Institute of Navigation.

Point Real-Time Kinematic Positioning

Y. Gao, M. Abdel-Salam, K. Chen and A. Wojciechowski
Department of Geomatics Engineering
2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4

Abstract. Autonomous point positioning was the original aim of GPS. However, due to errors caused by satellite ephemerides, satellite clock corrections, ionosphere, troposphere, multipath and noise, the accuracy of autonomous point positioning cannot be better than a couple of meters, even after Selective Availability (SA) was turned off.

Nowadays, with the emergence of Point Real-Time Kinematic (P-RTK) positioning, a lot of attention is again focused on standalone positioning. This method arose from the availability of precise ephemeris and satellite clock corrections. The concept behind this method is to make use of un-differenced code and carrier phase observations along with precise corrections to get the best accuracy out of GPS. The concepts, challenges, processing steps and results of the P-RTK system are presented in this paper.

Currently at the University of Calgary, a software system has been developed that achieves centimeter to decimeter level accuracy with a single GPS receiver. All components of P-RTK have been incorporated in this software, which is portable and user-friendly, containing many graphical interface tools for analyzing data.

Keywords: GPS, Precise Point Positioning, RTK

1 Introduction

Current carrier phase RTK positioning at centimeter level accuracy requires the combination of observations from a minimum of two GPS receivers. At least one of these serves as the base station with known coordinates, and the others serve as rover stations whose position coordinates are to be determined relative to the base station(s). Drawbacks of this approach include the practical constraints imposed by the requirement that simultaneous observations need to be made at the rover and base stations, and that the rover station should be in the vicinity of the base station(s), typically up to 20 kilometers.

The requirements for the deployment of base station(s) and the spatially limited operating range of the rover stations have increased the operational cost and logistical complexity in the field. As such, the full adoption of RTK technology has been limited in many applications.

This paper describes the concept of Point Real-Time Kinematic Positioning (P-RTK). This new RTK positioning approach is based on un-differenced carrier phase data using a single GPS receiver assisted by precise orbit and clock corrections. Since P-RTK has no requirement for the deployment of base station(s), it is a global RTK approach capable of providing greater solution consistency with increased operational flexibility. P-RTK will not only enrich the positioning options available to users, but also provide better time transfer capabilities and aid meteorological disciplines.

The remainder of the paper is organized into five sections. The P-RTK method is first described in Section 2. Various models for un-differenced carrier phase data processing are presented in Section 3. Challenges in point RTK are explored in Section 4. Numerical results are presented in Section 5 to assess the potential of the method and to identify problems that need to be solved in the future. Finally, several concluding remarks are given in Section 6.

2 Concept of Point RTK Positioning

Conventional Standard Point Positioning (SPP) that was initially designed for GPS is subject to the influence of all error sources. Major error sources include those introduced by broadcast orbits and clocks, as well as atmospheric effects. Since SPP is only able to provide position solutions with an accuracy level of several meters, it is not suitable as a positioning method for geodetic and surveying applications that require a positioning accuracy of several centimeters.

However, the situation has changed with the advent of precise orbit and clock correction products with centimeter level accuracy, since the two errors associated with broadcast orbits and clocks can be significantly reduced. Once orbit and clock errors are removed from GPS observations, much higher positioning accuracy can be expected even when a single GPS receiver is used. The

method that derives high precision positioning solutions by processing un-differenced carrier phase observations from a single GPS receiver assisted with precise orbit and clock corrections is called Precise Point Positioning (PPP) (Zumberge et al., 1997; Kouba and Heroux, 2000; Gao and Shen, 2002). The word "precise" is used here to distinguish it from the conventional SPP method.

PPP has received increased attention in the past several years within the GPS community due to its great operational flexibility and accuracy promise. Furthermore, if ambiguity resolution of un-differenced carrier phase observations becomes feasible, the PPP method will be ready to support the development of new RTK systems using a single receiver, without the need for base station(s). This is P-RTK, which is discussed in this paper.

PPP relies on the use of precise orbit and clock corrections. To date, there are many organizations, including the International GPS service (IGS), Natural Resources Canada (NRCan) and Jet Propulsion Laboratory (JPL), which offer precise data in both post-mission and real-time modes. IGS currently provides precise satellite orbit and clock correction data with different precisions and latencies: *predicted corrections* with accuracies of 25cm and 5ns, available twice daily, *rapid corrections* with accuracies of 5cm and 0.2ns, with a latency of 17 hours, and *final corrections* with accuracies of less than 5cm and 0.1ns, and with a latency of 13 days (IGS, 2000). JPL and NRCan provide real-time satellite orbit and clock corrections over the Internet for testing purposes.

3 Observation Models in Point RTK

The observation equations for code and carrier phase measurements on the Li frequency (i = 1, 2) are shown in Equations (1) and (2).

$$P(Li) = \rho + c(dt - dT) + d_{orb} + d_{trop} + d_{ion/Li} + d_{mult/P(Li)} + \varepsilon(P(Li)) \quad (1)$$

$$\Phi(Li) = \rho + c(dt - dT) + d_{orb} + d_{trop} - d_{ion/Li} + \lambda_i N_i + \lambda_i(\phi_r(t_0, Li) - \phi_s(t_0, Li)) + d_{mult/\Phi(Li)} + \varepsilon(\Phi(Li)) \quad (2)$$

where:

$P(Li)$	is the measured pseudorange on Li (m);
$\Phi(Li)$	is the measured carrier phase on Li (m);
ρ	is the true geometric range (m);
c	is the speed of light (m/s);
dt	is the satellite clock error (s);
dT	is the receiver clock error (s);
d_{orb}	is the satellite orbital error (m);
d_{trop}	is the tropospheric delay (m);
$d_{ion/Li}$	is the ionospheric delay on Li (m);
λ_i	is the wavelength on Li (m);
N_i	is the integer phase ambiguity on Li (cycle);
$\phi_r(t_0, Li)$	is the initial phase of the receiver oscillator;
$\phi_s(t_0, Li)$	is the initial phase of the satellite oscillator;
$d_{mult/P(Li)}$	is the multipath effect in the measured pseudorange on Li (m);
$d_{mult/\Phi(Li)}$	is the multipath effect in the measured carrier phase on Li (m) and
$\varepsilon(.)$	is the measurement noise (m).

Note that the initial phase of the receiver and satellite oscillators is commonly ignored in conventional double-difference RTK systems based on carrier phase. If it is combined with the integer phase components into a single term, Equation (2) can be rewritten as:

$$\Phi(Li) = \rho + c(dt - dT) + d_{orb} + d_{trop} - d_{ion/Li} + \lambda_i N_i' + d_{mult/\Phi(Li)} + \varepsilon(\Phi(Li)) \quad (3)$$

where N_i' is no longer an integer term.

In order to mitigate the ionospheric effect, which is the largest error source in GPS positioning after SA was turned off, ionosphere-free combinations are usually formed in dual-frequency receivers and have the following expressions:

$$P_{IF} = [f_1^2 \cdot P(L1) - f_2^2 \cdot P(L2)]/[f_1^2 - f_2^2] = \rho + c(dt - dT) + d_{orb} + d_{trop} + d_{mult/P(L1+L2)} + \varepsilon(P(L1+L2)) \quad (4)$$

$$\Phi_{IF} = [f_1^2 \cdot \Phi(L1) - f_2^2 \cdot \Phi(L2)]/[f_1^2 - f_2^2] = \rho + c(dt - dT) + d_{orb} + d_{trop} + \frac{cf_1 N_1' - cf_2 N_2'}{f_1^2 - f_2^2} + d_{mult/\Phi(L1+L2)} + \varepsilon(\Phi(L1+L2)) \quad (5)$$

Applying corrections from precise orbit and clock products to Equations (4) and (5) results in the following equations:

$$P_{IF} = \rho - cdT) + d_{trop} + d_{mult/P(L1+L2)} + \varepsilon\ (P(L1+L2)) \quad (6)$$

$$\Phi_{IF} = \rho - cdT + d_{trop} + \frac{cf_1N_1' - cf_2N_2'}{f_1^2 - f_2^2} + d_{mult/\Phi(L1+L2)} + \varepsilon\ (\Phi(L1+L2)) \quad (7)$$

Based on Equations (6) and (7), the unknowns to be estimated include the position coordinates, receiver clock offset, troposphere and a combined ambiguity term. The above model has been used in the observation model for PPP processing by Zumberg et al (1997) and Kouba and Heroux (2001).

The integer property of the ambiguity cannot be exploited based on Equations (6) and (7). Gao and Shen (2001) have proposed a new code combination to replace Equation (6) as follows:

$$P_{IF,Li} = 0.5[P(Li) + \Phi(Li)] = \rho - cdT + d_{trop} + 0.5\lambda_i N_i' + 0.5d_{mult/P(Li)} + 0.5\varepsilon\ (P(Li) + \Phi(Li)) \quad (8)$$

where i = 1 and 2.

Note that Equation (8) is still an ionosphere-free observable. A combination formed from Equations (7) and (8) yields an alternative observation model for PPP. Different from the model based on Equations (6) and (7), the new model is capable of estimating the ambiguities associated with L1 and L2 frequencies separately. This makes it possible to exploit the integer properties of both L1 and L2 ambiguities, which is essential for real-time kinematic positioning.

To facilitate high precision position determination using P-RTK, a number of unconventional error corrections have to be applied, including corrections for earth tides, satellite antenna offsets and phase wind up. These corrections are commonly ignored in conventional RTK positioning because they can be canceled out by the carrier phase double-differencing procedure that is implemented between satellites and receivers.

4 Challenges in Point RTK

There are three major challenges associated with P-RTK: error mitigation, carrier phase ambiguity convergence, and ambiguity resolution.

The ionosphere is a conventional error that causes the largest absolute error in GPS observations. Due to its dispersive nature, signals from satellites suffer different delays on L1 and L2. As such, they can be mitigated through standard ionosphere-free code and carrier phase combinations. The troposphere cannot be mitigated in this manner due to its non-dispersive nature. However, it can be modeled or estimated along with other parameters.

Unconventional error sources must also be taken into consideration in order to implement a P-RTK system. Many of these errors have been ignored in the past because they are irrelevant, negligible, or are canceled out through differencing. However, in the case of un-differenced code and carrier phase observations some of these errors do not cancel out and their sizes are relatively large, influencing the accuracy of the method.

These unconventional errors may be related to the un-differenced observations, the precise data or the standard GPS errors. Satellite antenna phase center, earth tide and ocean loading are examples of errors related to precise data. The satellite antenna phase center correction is necessary for Block II/IIA satellites because the phase centers and centers of mass of these satellites do not coincide. Earth tide and ocean loading models are necessary because errors associated with them can reach several decimeters. Similarly, a satellite phase windup correction is necessary since the error can reach half a cycle.

In most cases, carrier phase ambiguities are considered as float terms, often requiring long convergence times ranging from several tens of minutes to several hours. Since convergence is a crucial issue for real-time applications, long convergence times may prevent P-RTK from fulfilling the operational requirements of such applications. As such, fast ambiguity convergence methods and algorithms should be developed.

Integer ambiguities must be resolved to fully realize the accuracy of carrier phase observations. If integer values of the carrier phase ambiguities can be determined on-the-fly (OTF) over short time intervals, the P-RTK method will be able to support real-time kinematic positioning at centimeter level accuracy. Currently, this accuracy level is only feasible using double-difference RTK systems.

New ionospheric-free observation combination models are necessary because the conventional model, which forms ionospheric-free code and carrier phase combinations, does not allow for the exploitation of the integer property of the ambiguities.

Estimation of integer ambiguities is difficult due to the existence of non-zero initial phase offsets associated with the satellite and receiver oscillators and the level of residual errors that may exist in the system due to precise data and noise. Non-zero initial phase offsets are constant biases associated with the time and geometry of receiver-satellite lock, and are constant during an observed cycle slip free satellite arc (Gabor, 2000; Teunissen, 1997).

5 Numerical Results and Analysis

A software package called P^3 has been developed at the University of Calgary for precise point positioning. The software can be used to assess the performance of different data processing models as well as the influence of different error sources on positioning results.

Processing can be done in post-mission or in real-time, and the program can be run in either static or kinematic mode. Two point positioning modes are available: Single Point Positioning (SPP), which only makes use of code measurements, and Precise Point Positioning (PPP), which makes use of both code and phase measurements, as well as precise satellite orbit and clock corrections.

Processing in real-time requires real-time corrections. Currently, the developers of the software can obtain GPS•C data, which are real-time GPS wide-area corrections available from Natural Resources Canada (NRCan). GPS•C provides enhanced real-time positioning by applying corrections to user GPS receiver observations (GPS•C ICD, 2001).

5.1 Static Results

Shown in Figures 1 and 2 are static data processing results using final precise orbit and clock corrections from IGS, with accuracies of 5cm and 0.1ns, respectively. A one-day's data set collected on February 15, 2003, at the IGS station ALGO was used with a sampling interval of 30s. Precise satellite coordinates were available every 15min, and precise clock corrections every 5min. Due to the discrepancy in the sampling interval between the GPS observations and the precise data, the precise orbit and clock corrections were interpolated to 30s, equivalent to the sampling interval of the GPS observations. It should be noted that this may degrade the accuracy of the satellite orbit, and in particular the satellite clock corrections.

Positioning accuracy statistics are given in Table 1. The results indicate that the position solution converges to 2cm in 30min with a bias below 1cm.

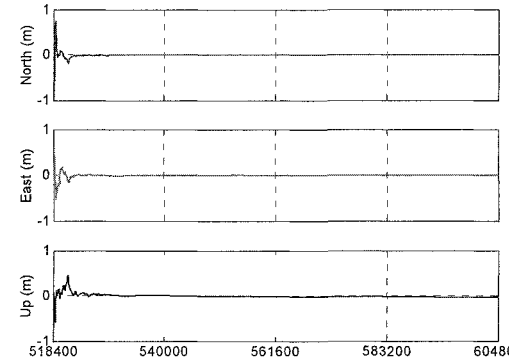


Fig. 1 Position Error Using IGS Final Products

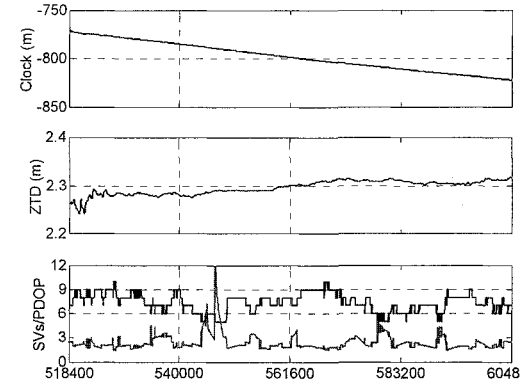


Fig. 2 Clock, Troposphere, PDOP and Number of Satellites Using IGS Final Products

Table 1. Static Positioning Accuracy Statistics Using IGS Final Products

	North	East	Up
Mean (m)	0.001	0.002	-0.006
RMS (m)	0.005	0.005	0.019
STD (m)	0.005	0.004	0.018

5.2 Kinematic Results

In this section, the same IGS data set was processed in kinematic mode. In addition, a two-hour data set (1 Hz) was processed in kinematic

mode using real time GPS•C corrections. GPS•C ephemeris and satellite clock corrections are accurate at the level of 20-25cm and 1-2ns, respectively (GPS•C ICD, 2001).

As mentioned above in Section 5.1, precise data had to be interpolated to 30s to be consistent with the sampling interval of the GPS observations. As such, two solutions using IGS final products are presented. Results of the first solution, using clock corrections interpolated at 30s, are shown in Figures 3 and 4, with statistics in Table 2. Results of the second solution, using the original 5min clock corrections, are shown in Figures 5 and 6, with statistics in Table 3.

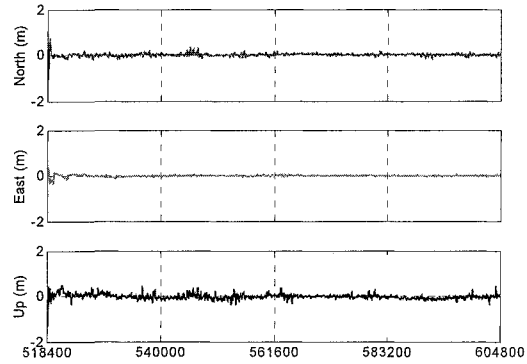


Fig. 3 Position Error Using IGS Final Products Interpolated at 30s

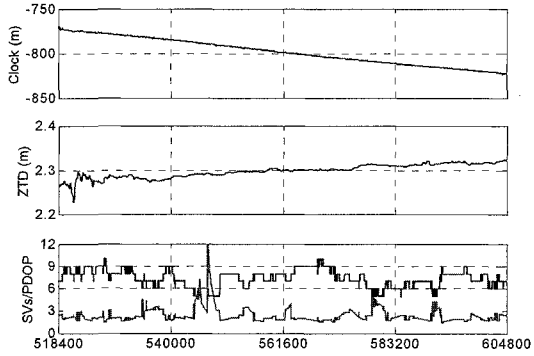


Fig. 4 Clock, Troposphere, PDOP and Number of Satellites Using IGS Final Products Interpolated at 30s

Table 2. Kinematic Positioning Accuracy Statistics Using IGS Final Products Interpolated at 30s

	North	East	Up
Mean (m)	0.014	0.001	-0.021
RMS (m)	0.054	0.023	0.098
STD (m)	0.052	0.023	0.096

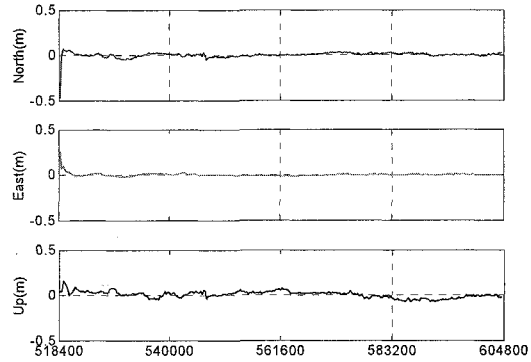


Fig. 5 Position Error Using Original IGS Final Products

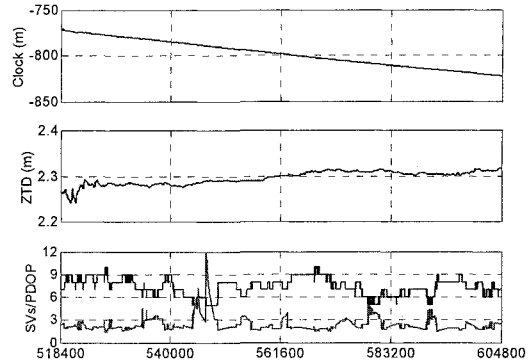


Fig. 6 Clock, Troposphere, PDOP and Number of Satellites Using Original IGS Final Products

Table 3. Kinematic Positioning Accuracy Statistics Using Original IGS Final Products

	North	East	Up
Mean (m)	0.004	0.000	-0.006
RMS (m)	0.016	0.009	0.036
STD (m)	0.016	0.009	0.035

Comparing Figure 3 and Table 2 with Figure 5 and Table 3 shows the degradation of results after an interpolation of clock corrections from 5min to 30s. Without interpolation the results converge very quickly, as shown in Figure 5. After convergence, the bias is less than 1cm, and the errors are below 5cm. The results reflect the accuracy of IGS final products. When using clock corrections interpolated to 30s, convergence occurs after about 15min, with about 5cm horizontal accuracy and 10cm vertical accuracy. Along with station coordinate parameters, estimation of the troposphere and receiver clock offsets are presented in Figures 4 and 6. These two products have a demand from meteorological and time transfer sectors.

Figures 7 and 8 show kinematic results using GPS•C real-time corrections for a two-hour data

set. A summary of the results is shown in Table 4. 10cm accuracy in the horizontal direction and 30cm accuracy in vertical direction is achievable with a convergence time of 30min. The results of IGS data and GPS•C data suggest that convergence depends on the accuracy of the precise data.

The results obtained with GPS•C corrections might be improved since the system is still in a test phase. With GPS•C real-time clock corrections at a two-second interval, no accuracy degradation associated with interpolation will occur. Moreover, GPS•C corrections can be widely used in real-time applications since they can be provided to users in Canada as well as to global users in real-time.

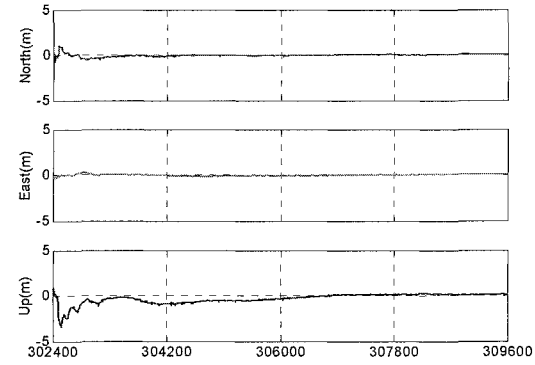


Fig. 7 Position Error Using GPS•C Real-Time Corrections

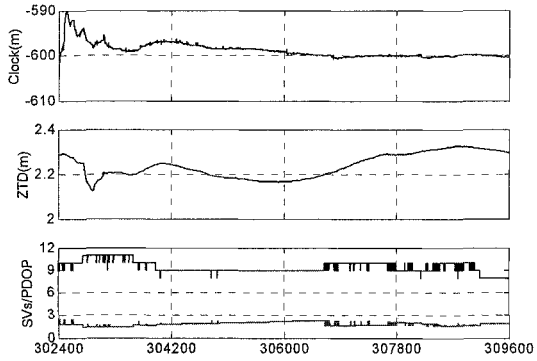


Fig. 8 Clock, Troposphere, PDOP and Number of Satellites Using GPS•C Real-Time Corrections

Table 4. Kinematic Positioning Accuracy Statistics Using GPS•C Real-Time Corrections

	North	East	Up
Mean (m)	0.040	0.002	-0.051
RMS (m)	0.062	0.063	0.280
STD (m)	0.048	0.063	0.275

6 Conclusions

The method of Point Real-Time Kinematic Positioning (P-RTK) has been described in this paper. Compared to the conventional RTK method, P-RTK can offer much greater operational flexibility to help in the full implementation of RTK technology in many applications in the future.

Test results have indicated that P-RTK positioning at centimeter to decimeter level accuracy is achievable. Using IGS final products, an accuracy of several centimeters has been achieved. Using real-time GPS•C data, an accuracy below 10cm is achievable in the horizontal components, and below 30cm in the vertical direction. Although the results are promising, several challenges still exist with respect to error mitigation, carrier phase ambiguity convergence, and ambiguity resolution.

Acknowledgement

Mr. Pierre Heroux from Natural Resources Canada is acknowledged for many valuable discussions. This research is partially supported by GEOIDE.

References

Gao, Y. and X. Shen (2002). "A New Method Of Carrier Phase Based Precise Point Positioning", *Navigation:* Journal of the Institute of Navigation, Vol. 49, No. 2.

IGS, (2000). The IGS-2000 Annual Report.

Kouba, J and P. Heroux (2000). "GPS Precise Point Positioning Using IGS Orbit Products", *GPS Solutions*, Vol.5, No.2, Fall.

Michael J. Gabor (2000) "Characteristics Of Satellite-Satellite Single Difference Widelane Fractional Carrier Phase Biases". *Proceedings of ION GPS 2000*, Salt Lake City, UT, 19-22 September.

Teunissen, P.J.G. and A. Kleusberg (Eds.) (1998). *GPS for Geodesy*. Second Edition, Springer-Verlag, New York.

Zumberge James. F., M. M. Watkins and F. H. Webb (1997). "Characteristics And Applications Of Precise GPS Clock Solutions Every 30 Seconds". *Navigation,* Journal of the Institute of Navigation. 44(4), 449-456 winters 1997-1998.

GPS*C ICD, 2001, Natural Resources Canada. URL: www.cdgps.com/e-test/cdgps_documents/%5B14%5D%20-%20ICD%20-%20GPSC%20-%202001-02-06.pdf

The Integration of GPS and Pseudolites for Bridge Monitoring

J. Barnes, C. Rizos, H-K. Lee
Satellite Navigation and Positioning Group, School of Surveying and Spatial Information Systems
University of New South Wales, Sydney, NSW 2052, Australia

G. W. Roberts, X. Meng, E. Cosser, A. H. Dodson
Institute of Engineering Surveying and Space Geodesy, The University of Nottingham, Nottingham, UK

Abstract. One application in structural engineering is measuring the movement of suspension bridges. Under unfavourable wind conditions or heavy traffic loads cable suspension bridges may move up to a few metres. Therefore, a positioning system used to monitor bridge movements can provide extremely valuable information. For such monitoring applications it is desirable for the positioning system to deliver equal precision in all position components, all the time.

GPS has the potential to deliver cm-level positioning accuracy, through use of the very precise carrier-phase measurement. However, the same level of accuracy cannot be guaranteed in all three-position components, twenty-four hours a day, in every situation.

Additionally, GPS signals are very weak and easily obstructed, especially in built-up urban environments. One solution to improving the GPS satellite geometry and the availability of ranging signals is to use ground-based transmitters of GPS-like signals (called "pseudolites"). A major advantage of pseudolite devices is the ability to control where they are located, improving a geometrically poor GPS constellation.

In this paper proof of concept for the use of pseudolites in the monitoring of suspension bridges is demonstrated through an experimental trial conducted on the Parsley Bay suspension footbridge in Sydney, Australia.

Keywords. Pseudolite, GPS augmentation, structural deformation

1 Introduction

The Global Positioning System (GPS) has been used for many years for the deformation monitoring of man-made structures such as bridges, dams and buildings, as well as geodetic applications, including the measurement of crustal motion, and the monitoring of ground subsidence and volcanic activity.

One application in structural engineering is measuring the movement of suspension bridges. Under unfavourable wind conditions or heavy traffic loads cable suspension bridges may move up to a few metres. If bridge movements can be precisely measured then bridge designers can compare real and predicted deformation. With continuous or periodic measuring of a bridge's movement, the structural integrity can be monitored and warn of potential failure. For the monitoring of man-made structures, and specifically suspension bridges, it is desirable for the measurement system to deliver equal precision in all position components, all of the time.

GPS has the potential to deliver cm-level positioning accuracy, through use of the very precise carrier-phase measurement. However, the accuracy, availability, reliability and integrity of the GPS position solutions are very dependent on the number and geometric distribution of the available satellites. Therefore, the positioning precision is not the same in all three components, and large variations in positioning precision can be expected during a 24-hour period. This situation becomes worse when the line-of-sight to GPS satellites becomes obstructed, such as in urban environments.

Also, GPS data below an elevation of approximately 15 degrees is usually of poorer quality and typically not used, due to difficulties associated with error modelling and susceptibility to being obstructed. This fact and the geometric distribution of the GPS satellite constellation results in 2 to 3 times worse accuracy in the height component, in comparison to horizontal components. Additionally, in mid and high latitude areas (>45 degrees) the accuracy in the North component is worse than the East component, due to the 55 degree inclination of GPS satellites (Meng et al., 2002).

One solution to improving the GPS satellite geometry and the availability of ranging signals is to use ground-based transmitters of GPS-like signals (called "pseudolites"). A major advantage of pseudolite (PL) devices is the ability to control where they are located, improving a geometrically poor GPS constellation. With four or more pseudolite devices, positioning independent of GPS is possible, with the advantage of complete control over the fixed ground based signal transmission.

The Satellite Navigation and Positioning (SNAP) group at The University of New South Wales (UNSW) has been actively conducting research into high precision combined kinematic GPS-PL for the past three years. This has led to the development of innovative carrier-phase single-epoch software that can process both GPS and PL data. Due to the comparatively small separation between PLs and user receivers, there are some challenging modelling issues, such as non-linearity, PL location errors, tropospheric delays, multipath and noise. These issues have been discussed in a series of papers by SNAP researchers (Barnes et al., 2002a; Wang et al., 2001; Wang et al., 2000; Dai et al., 2000). In addition, not all GPS receivers can track PL signals, and there are near-far signal strength issues. Because of these difficulties PLs are not yet a mainstream 'off-the-shelf' technology.

Deformation monitoring is one application where SNAP has explored the use of PLs (Barnes et al., 2002b; Dai et al., 2000). More recently the SNAP group, together with the IESSG at Nottingham University (UK), have conducted research into the use of PLs for the monitoring of suspension bridges. The IESSG have been interested in the monitoring of suspension bridges using GPS and accelerometers for the past few years, and their research is documented in such papers as Meng (2002), Meng et al. (2002), Dodson et al. (2001).

In this paper the results and analyses are presented of the augmentation of GPS with PLs for a trial conducted on a suspension footbridge in Sydney, Australia.

2 Parsley Bay Suspension Footbridge Trial

Parsley Bay is a narrow inlet on the east side of Sydney Harbour, and is a popular picnic and swimming location. In 1910 a cable suspension footbridge (Figure 1) was constructed across the bay to improve pedestrian access between the two shores of the bay. The bridge was built to the design of Edwin Sautelle, who was Town Clerk and Engineer of Vaucluse Council, and cost 500 pounds. The bridge towers, deck and handrails are

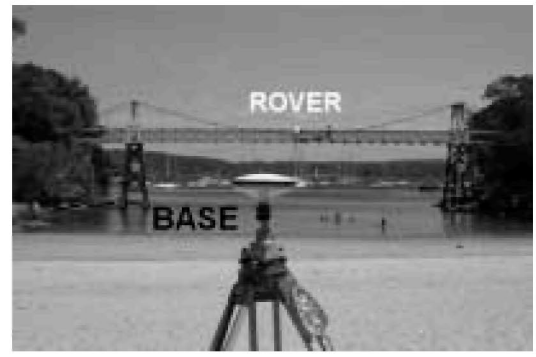


Fig. 1 Parsley Bay suspension footbridge trial.

constructed from wood, whilst the cabling and sides are steel. The bridge, now classified by the National Trust, is approximately 55 metres long and 2 metres wide.

On 16 January 2003 a trial was carried out on Parsley Bay Bridge to test the suitability of a PL-augmented GPS system for bridge monitoring. Two IntegriNautics PLs (PL12 and PL32) were used in the experiment, and their configuration is illustrated in Figure 2. Their locations were selected so that distances to both the reference receiver antenna (base) and the antenna mounted on the bridge (rover) were approximately equal, and with a clear line-of-sight. This was to ensure that the base and rover GPS receiver could track the PL signal with approximately the same signal strength. Additionally, the locations of the PLs were selected with a relatively unobstructed sky to allow the positions to be determined using GPS. The elevation angles of the PLs from the base antenna were close to zero, and at negative elevation angles from the rover antenna (PL12 –3.9 and PL32 –5.8 degrees).

Leica SR530 GPS receivers were used to survey the locations of the PLs with respect to the base

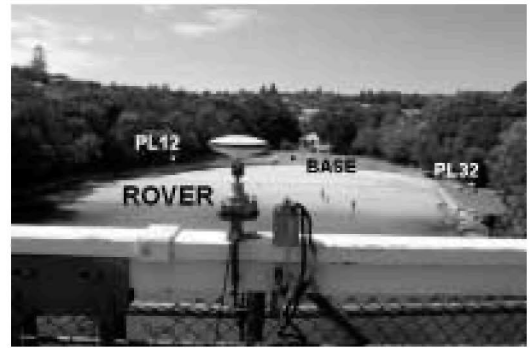


Fig. 2 Location of PLs, base and rover receiver.

antenna, by collecting approximately 30 minutes of data at a 1Hz rate. For the tracking of PL and GPS signals during the trial, dual frequency NovAtel Millennium (OEM3) GPS receivers were used. The NovAtel receiver allows the user to request particular PRN codes to be tracked, and this is a basic requirement for a PL-tracking GPS receiver. The PLs used Micropulse passive patch antennas, and were assigned GPS PRN codes 12 and 32. They operated in a pulsed mode with a 10% duty cycle, and their power was adjusted to give good signal-to-noise ratios (above 45-dBHz) at both the base and rover, but not too strong so as to jam the GPS signals. During the test approximately 2 hours and 20 minutes of data at a 2Hz rate were recorded from the NovAtel receivers, using laptop computers via a serial interface.

On six occasions during the trial movement of the footbridge was intentionally induced. This was achieved by 4 people rocking the bridge from side-to-side. Five of these occurred approximately between 1800-3300s and one at 7600s, from the start of the trial.

2.1 Geometry Analysis of GPS-only and PL Augmented GPS (GPS-PL)

It is important to investigate the rover geometry for GPS-only and pseudolite augmented GPS (GPS-PL) configurations during the trial period. There were between 4 and 7 satellites available above 15 degrees during the trial. Figure 3 shows the Dilution of Precision (DOP) in East, North and Vertical directions, together with the number of SVs/PLs available for GPS-only and GPS-PL configurations.

It is clear from Figure 3 that when the number of GPS satellites drops to 4, just after 7000 seconds, the satellite geometry is extremely poor, with all DOP values greater than 800! DOP values as large as this are likely to result in very poor positioning precision. It should be noted that the drop in satellites to 4 (above 15 degrees elevation) is purely due to a weak satellite constellation and not an obstructed sky-view. Also, a drop in satellite numbers from 6 to 5 has a major impact on the geometry, indicated by spikes in the DOP time series.

From Figure 3 it is clear that the pseudolites improve the GPS-only geometry greatly. When the number of GPS satellites drops to 4 and 5 in the GPS-PL configuration, the DOP values are not significantly affected, with all DOP values below 2.4.

To assess the improvement in the GPS constellation through augmentation with PLs the

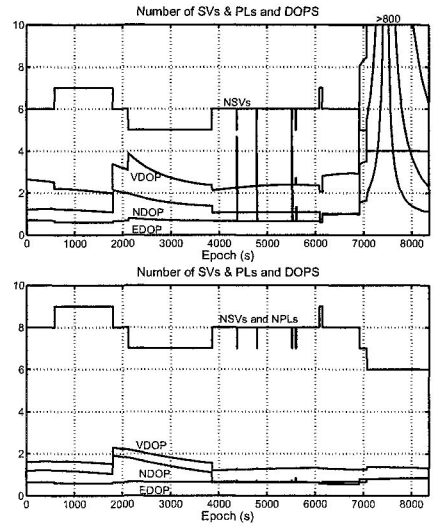


Fig 3. DOPs and NSVs for GPS-only (top) and GPS-PL (bottom) configurations.

mean, maximum and minimum DOP values were computed for both configurations and are given in Table 1. For this analysis, epochs with only 4 satellites were not used, since epochs with such high DOP values would not result in cm-level positioning and be rejected by most GPS processing software.

Overall, for both GPS-only and GPS-PL configurations, the best geometry is in East, then North, and lastly Vertical. For the GPS-PL configuration the mean DOP values are all less than 1. The improvement in the geometry (% Diff. in Table 1) when using PLs is greatest in the Vertical (42.6%), then North (20.9%) and least in East (17.2%).

Table 1. Summary of DOP values for GPS-only and GPS augmented with pseudolites, computed where there are more than 4 SVs.

		GPS-only	GPS-PL	% Diff.
EDOP	Max	3.51	0.83	
	Min	0.60	0.56	
	Mean	0.77	0.64	17.2
NDOP	Max	2.89	1.93	
	Min	0.74	0.54	
	Mean	1.29	1.02	20.9
VDOP	Max	8.46	2.30	
	Min	1.99	1.19	
	Mean	2.69	1.54	42.6

2.2 GPS and Pseudolite Data Processing

The baseline software developed by the SNAP group was used to process the GPS and PL data. For all the data processing a satellite elevation mask of 15 degrees was used, and the final solution type was L1 double-difference ambiguity-fixed. For PLs used in a static environment, over short distances, multipath is the dominant error and theoretically is a constant. For this bridge trial it was expected that the movements would be relatively small (less than one metre), and a constant pseudolite multipath bias was assumed. The multipath bias associated with the pseudolite must therefore be determined before using the PL carrier-phase data in the position computations. The methodology for this is described below.

Using only GPS data, an On-The-Fly (OTF) kinematic solution was first computed. The average rover position of the bridge was then computed using solutions with more than 4 satellites (due to the very poor DOP as indicated in section 2.1). Then, using the computed rover coordinates, carrier phase double-difference residuals were calculated for the GPS and PL data using SV14 as a reference satellite (assumed to have no long term biases). The double-difference residuals for the satellites were as expected, with small mean values (less than 3.5mm), indicating no significant biases. A typical satellite double-difference residual time series is given in Figure 4, for SV3 (elevation 45-15 degrees). The double-difference residual time series for the two pseudolites (in Figures 4) have significant constant biases, attributeable to multipath. The mean values of the time series, in millimetres, are –53.1 (PL12) and –70.8 (PL32). There are some time series fluctuations, in the GPS and GPS-PL time series, which are due to bridge movement and/or multipath.

Using the mean values, the PL data was corrected for the constant bias (due to multipath). Finally, in the data processing an OTF GPS-PL solution was computed. Where it was not possible to resolve ambiguities OTF, the average position of the bridge was used to resolve the carrier-phase ambiguities.

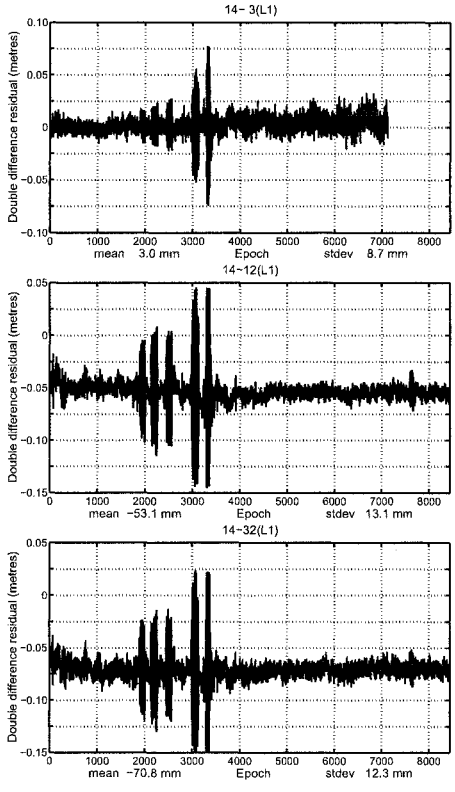


Fig. 4. Double difference residuals for SV3 (top), PL12 (middle) and PL32 (bottom).

2.3 Results of GPS-only and GPS-PL Positioning

Figure 5 shows the GPS-only positioning results in East, North and Vertical (mean-centred). From approximately epoch 7000s onwards, the position time series shows large position variation. This variation occurs when the number of GPS satellites drops to 4, resulting in very poor GPS geometry (see section 2.1). As mentioned in section 2, bridge movement was induced on six occasions, and one of which occured around epoch 7600s. However, it is difficult visually to separate the actual bridge movement from position variation induced by poor geometry. For the other 5 instances of induced movement (1800-3300s) the East and North time series show variation of several centimetres, and little variation in height.

Figure 6 shows the GPS-PL positioning results for East, North and Vertical. Compared with the GPS-only positioning result (Figure 5), there is not a large variation in the position time series from epoch 7000 s onwards. However, visually there is variation in the North position time series of up to 3cm around epoch 7600s (due to induced movement). Like the GPS-only solution, the induced bridge movement, around epochs 1800-3300s, is as large as 6.5cm and 10.0cm in the East and North components respectively.

Visually, it is also very clear that the GPS-PL position solution has less noise than the GPS-only solution. To demonstrate this fact standard deviations of the East, North and Vertical position

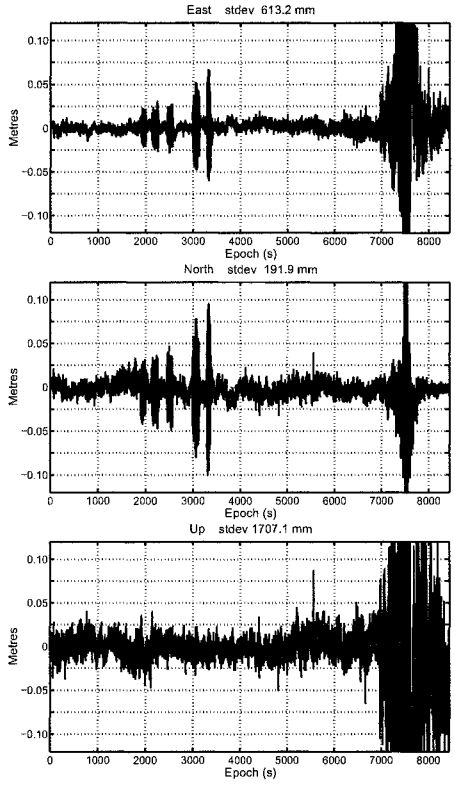


Fig. 5. GPS-only positioning result in East (top), North (middle) and Vertical (bottom).

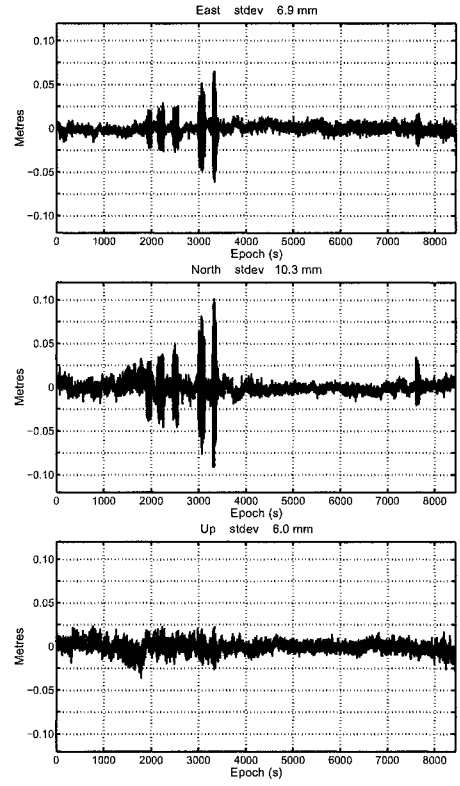


Fig 6. GPS-PL positioning result in East (top), North (middle) and Vertical (bottom).

time series are given in Table 2. These standard deviations are computed using data with more than 4 satellites, and not in periods of induced bridge movement. When GPS is augmented with PLs, the greatest improvement in the standard deviations is in the Vertical (58.7 %), then North (28.0%) and lastly East (27.1 %). The level of improvement is slightly better than that expected through geometry analysis (section 2.1). For the GPS-only solution the standard deviation in the Vertical (13.1mm) is approximately twice that of the horizontal components. Whereas for the GPS-PL solution the position component standard deviations are all less than 5.5mm, the Vertical precision is similar to the North component, and the East is slightly better. This result is the same as that expected through the geometry analysis referred to in section 2.1.

Table 2. Standard deviations of East, North and Vertical time series for GPS-only and GPS-PL, computed when there are greater than 4 SVs and no periods of induced bridge movment.

L1 single epoch solution type	Standard deviation (mm)		
	East	North	Vertical
GPS-only	5.3	6.9	13.1
GPS-PL	3.9	5.0	5.4
% Diff	27.1	28.0	58.7

2.5 Analysis of GPS-PL result

Does the GPS-PL solution time series correctly depict the actual movements of the bridge? For this it is useful to consider the orientation of the bridge with respect to the position components. From Figure 7 it can be seen that the centreline of the bridge runs approximately in a NE-SW direction. Therefore from the induced side-to-side bridge movement deviations in both horizontal components are expected (slightly more in North), and little in the Vertical. From Figure 6 this situation is clearly the case. Using the data period from approximately 2800-3300s, the maximum deviations in the positioning time series are 6.5, 10.0 and 2.6cm for East, North and Vertical respectively. For the Vertical component any movement is visually difficult to see above the

noise of the solution and the maximum horizontal deviation is 11.6cm.

Overall, the variations in the GPS-PL position time series are of the orders of magnitude expected.

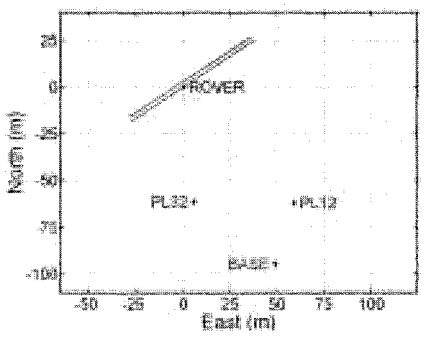


Fig. 7. Orientation of parsley bay bridge and relative locations of PLs and GPS receivers.

3. Concluding remarks

This paper has shown that there are situations when GPS cannot provide cm-level positioning accuracy twenty-four hours a day when using a 15 degree elevation mask, primarily due to poor constellation geometry.

By augmenting GPS with pseudolites it has been demonstrated that very poor geometry situations can be avoided. The introduction of the pseudolites in this trial improved the positioning precision by the greatest amount in the Vertical (58.7%), and by similar amounts for East (27.1%) and North (28%), resulting in similar positioning precision (sub-cm) for both horizontal and vertical components.

The pseudolite measurements suffered from constant biases due to multipath error, of which the largest was nearly 7.1cm. However the (almost) static nature of the trial allowed the multipath bias to be calibrated and removed from the pseudolite measurements.

All the induced bridge movements during the trial were visually apparent in the position time series and resulted in a maximum horizontal deviation of nearly 12cm. The vertical bridge movements were too small to be detected visually.

For a continuous deformation monitoring system there are situations when the GPS geometry can be very poor, and augmentation using pseudolites is one solution to the problem.

Acknowledgments

The authors would like to thank Woollahra Municipal Council for allowing the trial to be conducted at the Parsley Bay Bridge.

The collaborative research between SNAP and IESSG was aided through an Australian Research Council (ARC) International Linkage grant.

References

Barnes, J., J. Wang, C. Rizos, T. Nunan, C. Reid (2002a). The development of a GPS/pseudolite positioning system for vehicle tracking at BHP Billiton steelworks. In: *Proc of 15th International Technical Meeting of the Satellite Division of the US Inststitute of Navigation (ION-GPS 2002)*, 24-27 September, Portland, Oregan, pp. 1779-1789.

Barnes, J., J. Wang, C. Rizos, T. Tsujii (2002b). The performance of a pseudolite-based positioning system for deformation monitoring. In: *Proc of 2nd Symposium on Geodesy for Geotechnical & Structural Applications*, 21-24 May, Berlin, Germany, pp. 326-337.

Dai, L., J. Zhang, C. Rizos, S. Han, J. Wang (2000). GPS and pseudolite integration for deformation monitoring applications. In: *Proc of 13th International Technical Meeting of the Satellite Division of the US Institute of Navigation (ION-GPS 2000)*, 19-22 September, Salt Lake City, Utah, pp. 1-8.

Dodson, A.H., X. Meng, G. Roberts (2001). Adaptive method for multipath mitigation and its applications for structural deflection monitoring. In: *Proc of International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation (KIS2001)*, 5-8 June 2001, Banff, Alberta, Canada, pp. 101-110.

Meng, X. (2002). *Real-Time Deformation Monitoring of Bridges Using GPS/Accelerometers*. PhD thesis, The University of Nottingham, Nottingham, UK, 239 pp.

Meng, X., G.W. Roberts, A.H. Dodson, E. Cosser, C. Noakes (2002). Simulation of the effects of introducing pseudolite data into bridge deflection monitoring data. In: *Proc of 2nd Symposium on Geodesy for Geotechnical and Structural Engineering*, 21-24 May 2002, Berlin, pp. 372-381.

Wang, J., T. Tsujii, C. Rizos, L. Dai, M. Moore (2001). GPS and pseudo-satellites integration for precise positioning. *Geomatics Research Australasia*, 74, pp. 103-117.

Wang, J., T. Tsujii, C. Rizos, L. Dai, M. Moore (2000). Integrating GPS and pseudolite signals for position and attitude determination: Theoretical analysis and experiment results. In: *Proc 13th International Technical Meeting of the Satellite Division of the US Institute of Navigation*, 19-22 September, Salt Lake City, Utah, pp. 2252-2262.

HYBRID SENSOR SYSTEM FOR BRIDGE DEFORMATION MONITORING: INTERFACING WITH STRUCTURAL ENGINEERS

Xiaolin Meng, Alan Dodson, Gethin Roberts, Emily Cosser
Institute of Engineering Surveying and Space Geodesy (IESSG),
The University of Nottingham, NG7 2RD, UK

Abstract. Analytical finite element (FE) modelling, modal testing and FE model updating are approaches which are employed to detect structural dynamic characteristics in the mechanical and aerospace engineering communities. Their application to civil engineering structures is more recent. To successfully apply these approaches, field data acquisition plays a very important role in validating and updating the analytical model.

As one of the feasible data collection tools for modal testing, GPS has been employed with other sensors by the IESSG at The University of Nottingham to monitor large structural deformations. While demonstrating promising advantages in positioning precision and flexibility in instrumentation installation, the requirements from the structural engineering perspective need to be well defined before this technology can be fully implemented.

In this paper prototypes of hybrid sensor systems coupled with a computational FE model are introduced. Data collected from a test bed bridge are processed and analysed to demonstrate the feasibility of proposed systems. The results reveal that it is possible to achieve 3D millimetre positioning precision for detecting high dynamic structural deformations, and obtaining a highly accurate FE model for the purpose of structural health monitoring (SHM).

Keywords. GPS, Finite Element (FE) Model, Structural Deformation Monitoring

1 Introduction

A new science of structural health monitoring has emerged to assist the monitoring of structural integrity in recent years. One of the major tasks in SHM is the determination of structural dynamic models. Based on the parameters identified such as natural vibration frequencies, mode shapes and damping ratios, structural faults and fault locations can be detected through collecting precise field testing data and updating an analytical model such as a finite element (FE) model.

Two different methods are available to experimentally identify the structural dynamics: forced vibration testing (FVT) and ambient vibration testing (AVT). With FVT, known weight shakers are used to excite certain vibration modes. The output responses are measured by sensors such as accelerometers and more recently by GPS. The recorded results are then used to update analytical FE models. However, it is very difficult to find a device or method that can excite and detect the structural vibration frequencies of less than 2 Hz (http://www.rcidynamics.ch/), which are evident for large suspension bridges. For instance, the vertical vibration frequency of the Humber Bridge in the UK is 0.116 Hz (Brown et al. (1999)).

AVT is actually an output-only approach. It represents many real world scenarios where only response data, such as 3D geometric deformations or forces on the structures are available but the actual loading conditions (inputs) are totally or partially unknown. Typical advantages of AVT are that no expensive dynamic force generator is required, tests can be conducted without interrupting traffic, and it can be used for the bridges with an overall dimension exceeding 100m. However, compared with more mature FVT approach, AVT is still in the research phase.

The feasibility of GPS-based large structural deformation monitoring has been demonstrated in many GPS engineering applications in recent years (Fujino et al. (2000); Roberts et al. (1999)). In the presence of sufficient satellites with strong geometry, GPS can provide continuous, all-weather, automated and highly accurate measurements. The displacements measured with GPS can be compared with numerically predicted values from an established FE model either to update the model or to detect any potential structural problems.

The defects of GPS positioning compared to more traditional techniques for such applications have been investigated by Meng (2002). For instance, GPS signal availability is always problematic due to

surrounding obstructions, such as dense bridge cables, high rising bridge towers and buildings, etc. Also, due to the 55° inclination, satellite availability depends on geographical latitude and GPS satellites are not uniformly distributed around the globe. As a consequence, the positioning precision of the two horizontal components will not be at the same level, and it will cause misinterpretation of the actual structural deformations (Meng et al. (2003b)). Multipath, residual tropospheric delay and cycle slips are other major errors that need to be addressed when GPS is used for structural deformation monitoring. Slow sampling rate greatly limits the efficiency of a GPS alone system for this purpose. For example, the currently available 20 Hz sampling rate is insufficient to monitor bridge deformations with vibration frequencies higher than 10 Hz. Ways of improving the signal availability and augmenting satellite geometry, as well as how to increase the overall sampling rate of a monitoring system are the research focus. The correct extraction of structural dynamics from field testing data is another important issue from the perspective of structural engineers.

A hybrid deformation system, consisting of dual frequency GPS receivers and triaxial accelerometers, has been proposed initially by the researchers at the IESSG of The University of Nottingham (Dodson et al. (2001); Roberts et al. (2000a); Roberts et al. (2000b)). Hardware and software have been developed and validated in a series of bridge trials in the UK (Meng (2002); Roberts et al. (2000a)). With this hybrid system, dramatic improvements in productivity, system reliability and overall performance were achieved.

However, the data quality of this hybrid system largely relies on the quality of raw GPS measurements. The relative displacements from a triaxial accelerometer can only bridge very short gaps due to the drifts caused by the instrumental biases and scale factor errors. For example, the vertical position error can accumulate to 5m within 40 seconds using the force measurements collected by a Kistler triaxial accelerometer. Assuming that multipath and other systematic and random errors could be effectively modelled and mitigated, and that the sampling rate can be increased to track high structural dynamics, the defect caused by the GPS orbit design cannot be overcome by this system.

Pseudolites are ground based transmitters of GPS-like code and carrier phase signals on either L1 or L2, and can be used to enhance GPS by providing increased accuracy, integrity and availability (Elrod and Dierendonck (1996)). They are ideal for augmenting GPS constellation for precise deformation monitoring of bridges with limited movements (Dai et al. (2001); Barnes et al (2003)). In these applications, properly installed pseudolites can be used as another ranging source, which can improve the dilution of precision (DOP) and hence the overall quality of the solution. However, error sources relevant to pseudolites based positioning such as tropospheric delay and multipath, as well as time synchronisation and near-far problem need to be further investigated.

In this paper, a suspension footbridge is used as a test bed for the proposed research. Predicted results from an FE model developed by project partners at Cranfield University, are compared with the structural dynamics extracted from the field testing data. Discussions are then made regarding the selection of the optimal sensor locations, and signal extraction for updating the analytical model. The research involved in these aspects form the interface between geodesists and structural engineers. Aimed at monitoring structural health, different hybrid sensor integration approaches are introduced. The authors demonstrate that by using the proposed sensor systems it is possible to achieve 3D millimetre positioning precision for continuous structural deformation monitoring.

2 A Test Bed for Structural Deformation Monitoring

The research presented in this paper is funded by the Engineering and Physical Sciences Research Council (EPSRC) in the UK. The overall objective of the project is the creation of a system employing advanced computational tools such as FE modelling, coupled with GPS, pseudolites and triaxial accelerometer sensors, to remotely monitor the health of operational bridges without on-site inspection.

The Wilford suspension bridge over the River Trent in Nottingham was chosen as a test bed for this project because of its high dynamics under normal loading conditions and convenience to access. It is a 66m long and 3.5m wide footbridge. This bridge consists of two sets of suspension cables restrained by two masonry anchorages (Figure 1). A steel deck covered by a floor of wooden slats forms the bridge deck. Three water/gas pipes are laid underneath the wooden deck for supplying water and gas to the local residents.

In the last four years, a number of trials were conducted on this bridge to test the algorithms and hardware. Different instruments were installed on

the bridge, which included dual/single frequency GPS receivers, triaxial accelerometers, pseudolites, surveying total stations and meteorological meters. A complete procedure for bridge deflection monitoring was established, and successfully applied to the deformation monitoring of other bridges such as the London Millennium Bridge (Roberts et al. (2000b)).

As shown in Figure 1, two closely setup reference stations were used to mitigate multipath (Roberts et al. (2001)). To detect bridge dynamics, the number and locations of the sensors are crucial. The characterisation of the dynamic behaviour of a structure, or the identification of the presence of structural faults, is possible only under certain conditions. This requires having a minimum amount of information collected or a minimum number of sensors installed. Therefore, the sensors should be strategically placed in order to provide a clear "image" of the structural behaviour. However, due to economical issues, usually only a limited number of sensors are placed on some accessible locations on the real structures. Many approaches can be employed to find optimal sensor locations. Figure 2 shows the sensor locations determined by an optimal algorithm, EFfective Independence-Driving-Point Residue (EFI-DPR) (Meng et al. (2003a)). Details of this approach are beyond the scope of this paper.

Fig. 1 The Wilford Bridge and Dual Reference Stations

The construction of a finite element model was undertaken using the SAFESA™ method, an FE algorithm developed by Cranfield University. The 3D FE model developed for the Wilford bridge is shown in Figure 3. With the recommended instrument layout, sufficient information about bridge dynamics can be obtained.

Table 1 lists the relevant natural vibration frequencies estimated by the established FE model.

While the first natural frequencies extracted from the accelerometer and GPS measurements are identical using Fast Fourier Transform (FFT), the frequencies estimated by the analytical FE mode do not match well. This could be partially due to the fact that there were no detailed material and design descriptions about this old bridge. At this stage, the initial FE model is not accurate enough for health monitoring purposes. Therefore updating the FE model with precise field measurements is required. GPS/FEM approach could be then used to monitor the changes of structures and detect possible structural faults, through collaborative research between two universities with complementary expertise.

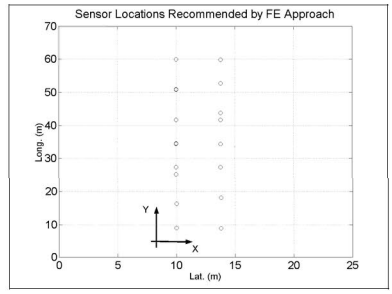


Fig. 2 Rover Station Layout.

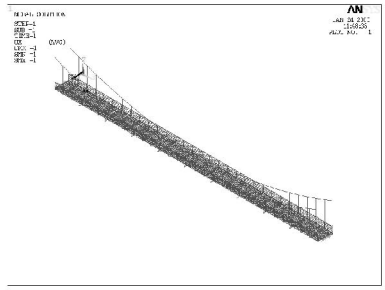


Fig. 3 3D FE Model for the Wilford Bridge.

Table 1 Estimated Natural Frequencies by FE Model.

Frequencies	
1	1.3797
2	2.6949
3	4.4851
4	6.5006
5	8.6137
6	10.72

3 A Hybrid Sensor System

Over the last couple of years, different hybrid sensor systems have been investigated. Initially, the research focus was on the expansion of the measurable sampling rate to detect higher structural dynamics. The research then evolved to develop a system which can be used to identify both small and large dynamic deformations with millimetre 3D positioning precision. Both systems are compared in

the following sections, using sample data sets collected from bridge trials.

3.1 Integration of GPS/Accelerometers

The vertical vibration frequencies of the Wilford bridge could reach 20 Hz which are out of the detectable range of a GPS receiver gathering data at a rate of 20 Hz (see Table 1 and also Figure 4). Figure 4 shows 3D vibration frequencies in a bridge coordinate system (BCS), a defined computational reference system with its longitudinal axis along the bridge major axis, lateral axis perpendicular to the major axis and vertical axis coinciding with the plumb line. The frequencies are extracted from a sample data set collected by a Kistler triaxial accelerometer during a GPS and accelerometer based trial on the Wilford bridge on 20 June 2002 (Figure 5). The sampling rate in this trial was 100 Hz. Five obvious vertical movements in Figure 5 were excited by force. From Figure 4, it can also be seen that more vibration frequencies were evident in the lateral direction of the bridge due to strong lateral wind loading and the external force applied on the bridge (people jumping on the bridge when tests were made), which excited more mode shapes than in the vertical or longitudinal directions.

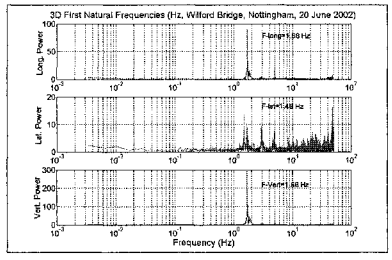


Fig. 4 3D Frequencies for the Wilford Bridge.

Figure 6 shows 3D mean centered displacements estimated from GPS measurements at the same time when the acceleration data were collected as illustrated in Figure 5. It is difficult to see the clear vertical movements caused by excitations, due to the high level of multipath and other receiver related random noises. Most vibration frequencies which are still measurable by GPS are buried undetectably by the noise spectra, as shown by Figure 7.

To increase the sampling rate and also improve positioning precision, a hybrid system consisting of dual frequency GPS receivers and triaxial accelerometers was proposed. The algorithms and data processing procedures are documented in Meng (2002). Figure 8 shows an integrated vertical position result for a period of five minutes, estimated from 10 Hz GPS positions and 100 Hz relative displacements by double integration of the vertical accelerations. It can be found that through GPS and accelerometer data integration the sampling rate can be expanded to the same level as an accelerometer, and more vibration frequencies can be detected by the FFT algorithm (see the low graph in Figure 8).

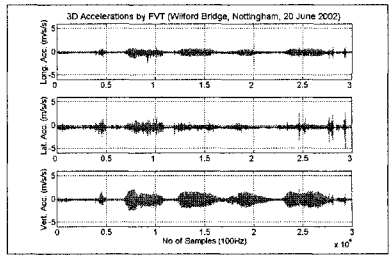


Fig. 5 3D Accelerations Excited by Force.

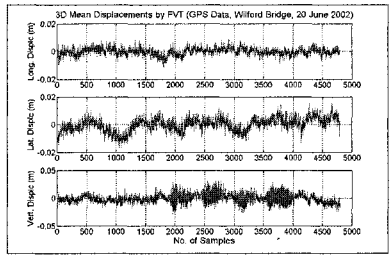


Fig. 6 3D Mean Displacements by GPS.

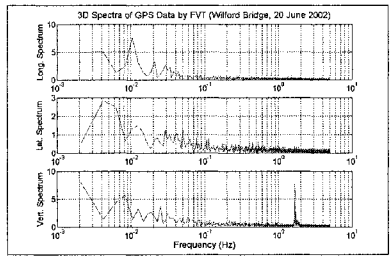


Fig. 7 3D Spectra of GPS Data.

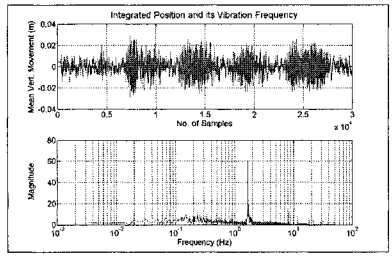


Fig. 8 Integrated Positions and Frequency Distribution.

3.2 Integration of GPS and Pseudolites

While the initially proposed sensor system could increase the data sampling rate and bridge the short

time gaps of GPS positioning, it cannot improve the data quality caused by poor satellite geometry and signal availability.

One effective way to solve the problem is to introduce several ground-based pseudolites to form stronger geometry of signal transmitters. Simulated results reveal that it is possible to achieve uniform millimetre level 3D positioning accuracy. Selection of optimal pseudolite locations is critical. Criteria for this have been investigated by Meng (2002) using a DOP simulator, and further validated by the actual pseudolite configurations.

Figure 9 shows the instrument layout for the bridge deformation monitoring trial jointly conducted with The University of New South Wales in Australia on 16th October 2002. In this trial, three Integrinautics' IN200 pseudolite generators were setup on the northern side of the Wilford bridge to augment the GPS satellite geometry. Pseudolite PL1 was setup on a tripod on the west bank of the River Trent. PL2 and PL3 were set up on the east bank. Two Allstar single frequency receivers capable of receiving GPS and pseudolite signals were used as reference and rover stations. It should be noted that during the data collection the sampling rate was set to 1 Hz, which can only be used to assess the improvement in the positioning solutions due to the augmented system.

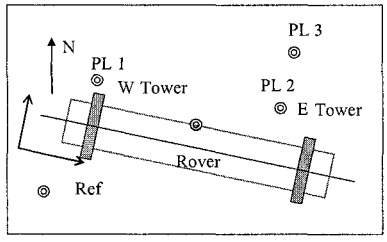


Fig. 9 Pseudolite Setup for a Bridge Trial.

Since there are only very small relative movements between pseudolites and receivers, multipath caused by the surroundings was assumed as to be constant bias. This can be identified by phase measurement double differencing. Multipath can then be removed from the raw measurements. Figure 10 shows the phase double differencing residuals between the base satellite PRN5 and PL1 (designated as PRN12). The mean value of these residuals is 51.1 mm. This could be due to signal obstruction from the arches of the west bridge anchorage. After removal of this constant bias from the phase measurements, using the approach recommended by Barnes et al. (2003), the mean residuals are reduced to -0.3 mm.

Using the multipath corrected phase measurements and a GPS/pseudolite post-processing software package developed by the SNAP group in The University of New South Wales, each epoch positions have been solved.

Figures 11, 12 and 13 show the mean centred easting, northing and vertical coordinates for the pseudolite augmented system. Comparing the standard deviations between GPS and the augmented systems, improvements at 43.4%, 13.3% and 34.6% were achieved in easting, northing and vertical component respectively. Also the vertical positioning precision is higher than that of the northern component with augmented system. These results confirm the simulated solution calculated by Meng (2002). To improve the positioning accuracy in the northern component to the same level as the easting or vertical component, some pseudolites should be set above the horizon of the receivers. However, for this particular observation environment, it is not feasible to find such locations.

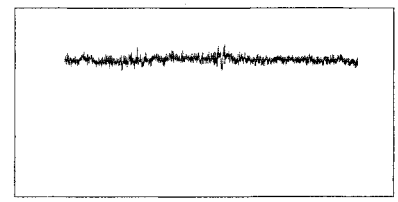


Fig. 10 Phase DD before Removal of Multipath.

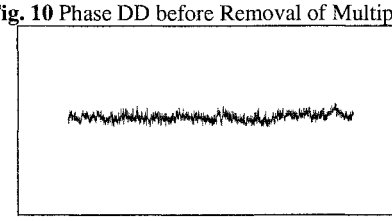


Fig. 11 Mean Centred Easting Positions from GPS/PLs

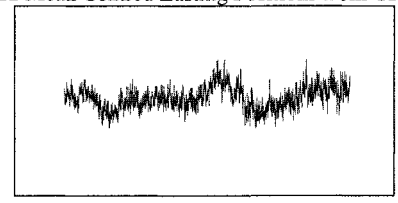


Fig. 12 Mean Centred Northing Positions from GPS/PLs

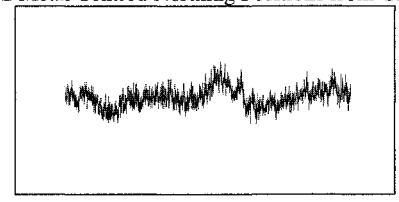


Fig. 13 Mean Centred Vertical Position from GPS/PLs.

Conclusions

Aimed at detecting structural dynamics for health monitoring using GPS, accelerometers and pseudolites, this paper has discussed the interface between structural engineers and geodesists. To achieve this, a test bed bridge for the project was introduced. An analytical FE model was developed for this bridge. Initially predicted results by the FE model were compared with structural dynamics extracted from field testing data. The needs to update this model were emphasised. Two sensor systems, for collecting data, were introduced with particular focus on the signal processing for extracting structural dynamics. The authors pointed out that with a hybrid sensor system consisting of GPS and accelerometers, the sampling rate can be increased to the same rate as the accelerometer and short time millimetre positioning precision can be assured given that there is a strong satellite geometry, and sufficient number of satellites. The authors then proposed a hybrid system comprising of ground based pseudolites and GPS satellites for improving the transmitter geometry. According to the data processing for an actual bridge trial, 3D positioning precision can reach the millimetre level. In trials with proposed system, an improvement in precision of up to 43% was achieved. This paper demonstrated through real examples that the integration of GPS, accelerometers and pseudolites can overcome the defects in GPS satellite geometry, signal availability, positioning accuracy and sampling rate.

Acknowledgements

The authors would like to acknowledge EPSRC in the UK for providing the funding for this project (Ref: GR/R28218/01). The supportive research of the Cranfield University is gratefully acknowledged. The Australian Research Council International Linkage programme supported Dr Barnes's travel to Nottingham to conduct the GPS/pseudolite based bridge trial and his assistance in the data processing is highly appreciated.

References

Barnes, J., C. Rizos, R. Warnant, X. Meng, E. Cosser, A.H. Dodson, G.W. Roberts (2003). The Monitoring of Bridge Movements using GPS and Pseudolites. In: *Proc of 11th International Symposium on Deformation Measurements, International Federation Surveyors (FIG), Commission 6 - Engineering Surveys, Working Group 6.1*, 25-28 May, Santorini, Greece.

Brown, C.J., R. Karuna, V. Ashkenazi, G. Roberts, R.A. Evans (1999). Monitoring of Structures Using the Global Positioning System. In: *Proc Instn Civ. Engrs Structs & Bldgs*, February, pp. 97-105.

Dai L., J. Wang, C. Rizos and S Han (2001). Pseudo-satellites Applications in Deformation Monitoring. *GPS Solutions*, 5(3), pp. 80-87.

Dodson, A.H., X. Meng and G. Roberts (2001). Adaptive Method for Multipath Mitigation and Its Applications for Structural Deflection Monitoring. In: *Proc of International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation (KIS2001)*, Banff, Alberta, Canada, 5-8 June, pp. 101-110.

Elrod, B.D. and A.J.V. Dierendonck (1996). Pseudolite. In *Global Positioning System: Theory and Applications*, edited by B.W. Parkinson, J.J. Spilker, P. Axelrad, and P. Enge, American Institute of Astronautics (AIAA), Inc., Washington, D.C., pp. 51-79.

Fujino, Y., M. Murata, S. Okano, M. Takeguchi (2000). Monitoring System of the Akashi Kaikyo Bridge and Displacement Measurement Using GPS. In: *Proc of Nondestructive Evaluation of Highways, Utilities, and Pipelines IV, Proceedings of SPIE*, edited by A.E. Aktan, and S.R. Gosselin, pp.229-235.

Meng, X. (2002). Real-Time Deformation Monitoring of Bridges Using GPS/Accelerometers. PhD thesis, IESSG, The University of Nottingham, Nottingham, UK.

Meng, X., M. Meo, G.W. Roberts, A.H. Dodson, E. Cosser, E. Luliano, A. Morris (2003a). Validating GPS Based Bridge Deformation Monitoring with Finite Element Model. In: *Proc of GNSS 2003, The European Navigation Conference*, 22-25 April, Graz, Austria.

Meng, X., G.W. Roberts, E. Cosser, A.H. Dodson, J. Barnes, C. Rizos (2003b). Real-time Bridge Deflection and Vibration Monitoring Using an Integrated GPS/Accelerometer/Pseudolite System. In: *Proc of 11th International Symposium on Deformation Measurements, International Federation Surveyors (FIG), Commission 6 - Engineering Surveys, Working Group 6.1*, 25-28 May, Santorini, Greece.

Roberts, G.W., A.H. Dodson, and V. Ashkenazi (1999). Twist and Deflection: Monitoring Motion of Humber Bridge, *GPS World, 10* (10), pp. 24-34.

Roberts, G.W., A.H. Dodson, C.J. Brown, R. Karuna, E. Evans (2000a). Monitoring the Height Deflections of the Humber Bridge by GPS, GLONASS and Finite Element Modelling. In *Proc of IAG Symposia Geodesy Beyond 2000*, edited by Schwarz, Springer-Verlag, Berlin, pp. 355-360.

Roberts, G.W., X. Meng, C.J. Brown (2000b). The Use of GPS for the Measurement of Bridge Movements: A Viability Study. The University of Nottingham and Brunel University.

Roberts, G.W., X. Meng, A.H. Dodson (2001). Data Processing and Multipath Mitigation for GPS/Accelerometer Based Hybrid Structural Deflection Monitoring System. In: *Proc of ION GPS 2001, 14th Int. Tech. Meeting of the Sat. Div. of the U.S. Inst. of Navigation*, 11-14 September, Salt Lake City, USA, pp. 473-481.

A Fuzzy System for the Assessment of Landslide Monitoring Data

M. Haberler
Institute of Geodesy and Geophysics, Department of Applied and Engineering Geodesy, Vienna University of Technology, Gusshausstrasse 27-29, 1040 Vienna, Austria

Abstract. In many cases landslide areas can be divided into several blocks, which are moving with different velocities in different directions. So, if we are able to detect block boundaries, landslide monitoring can be performed more efficiently. The information about the relative movement of these blocks is a very important indicator for future movement behavior, if monitored very precisely at the block boundaries with continuously measuring geotechnical sensors.

To detect the boundaries of the blocks the following algorithm is used: the displacement vectors of the observed points (out of a geodetic deformation analysis) will be analysed by an affine coordinate transformation. The assignment of the observed points to the different blocks is done by an iterative algorithm; the thresholds for the several steps of the algorithm are calculated by a fuzzy system. The input parameters for this fuzzy system are e.g. the standard deviations of the transformations and strain parameters calculated from these transformation steps. Finally, an example for application of this fuzzy system will be given.

Keywords. Landslide monitoring, fuzzy system, affine coordinate transformation

1 Introduction

Landslides are one of the major types of natural hazards worldwide. Every year thousands of people are injured or even killed. Additionally, many buildings and infrastructure like railroads and traffic networks are destroyed. E.g., only in the US, 25 to 50 people are killed per year and the economic damage reaches US$ 2 billion [8]. So, there is an urgent need for a suitable monitoring and alert system. The project OASYS (Integrated Optimisation of Landslide Alert Systems) was started to fulfil these requirements by a multi-disciplinary approach [3].

OASYS consists of several steps:
- At a regional scale, potential landslide areas have to be identified. This is done by satellite images (remote sensing data, INSAR), topographic and geological maps, historical reports,...
- At a local scale, for each of these landslide areas a monitoring system is installed. A geodetic network is observed in several epochs and based on these measurements the block boundaries can be identified.
- In a next step, high precision geotechnical sensors are installed across the block boundaries to permanently observe the relative movement of the blocks.
- A knowledge-based system is used to analyse the collected data and to support the decision making authorities in case of danger.

This paper deals with one part of the project OASYS, the detection of the block boundaries based on geodetic deformation measurements.

2 Algorithm for block detection

We assume that the geodetic networks have been measured periodically and the coordinates and displacement vectors for all observed points were calculated out of a geodetic deformation analysis. The block detection algorithm is the necessary step to proceed from the measurements of the geodetic network to the following installation of the high precision geotechnical sensors in the area under investigation (see section 1).

The block detection algorithm (see figure 1) starts with the identification of a minimal block of four points (explanation see section 2.1); in the following iterative procedure other neighbouring points are added to the block until it is 'complete'. This assessment is done by a fuzzy system. If the last point does not fit the block, it has to be removed, and the algorithm starts again, identifying

another minimal block out of the remaining observation points. The practical implementation of the algorithm is described in section 4.

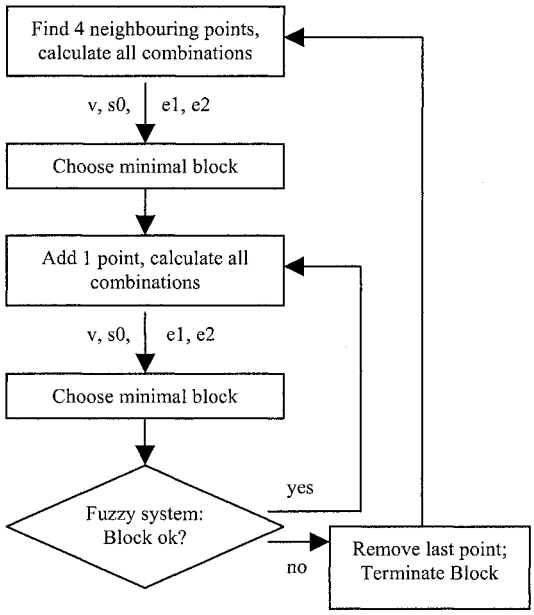


Fig. 1 A coarse scheme of the analysis algorithm implemented in MATLAB. Indicators for the assessment are the standard deviation s0, the residuals v and the semi-axes of the strain ellipse e1 and e2.

2.1 Affine coordinate transformation

The block detection algorithm is based on the displacement vectors of the observed points. The idea is to use an over-determined affine coordinate transformation to assess the movement of the points between two subsequent epochs of measurements. This means that the coordinates of the points of epoch n are mapped onto the coordinates of the same points of epoch n+1. A group of points moving in the same direction (assuming that they are lying on one common block) will give small residuals and a small standard deviation within an over-determined affine transformation. In case points of different blocks were considered simultaneously the standard deviation will be significantly larger.
In two dimensions, the approach reads as follows:

$$y_{n+1} = ay_n + bx_n + c$$
$$x_{n+1} = dy_n + ex_n + f$$

where y_n, x_n...coordinates of epoch n
y_{n+1}, x_{n+1}...coordinates of epoch n+1
a,..,f...transformation parameters

To solve this equation system, three identical points are necessary; for an over-determined solution at least 4 points have to be used. So the algorithm starts with a minimum block size of 4 points.

The six transformation parameters (a,..,f) can be interpreted as two translations (t_y, t_x), two rotations (w_y, w_x) and two scale parameters (m_y, m_x):

$$a = m_y \cos w_y \quad b = m_x \sin w_x \quad c = t_y$$
$$d = -m_y \sin w_y \quad e = m_x \cos w_x \quad f = t_x$$

The six-parameter approach was chosen because of the special properties of landslides. The movement of a block can be described by a translational part and a rotation/distortion [1]. In most cases the distortion in the direction of the movement is larger than in other directions. So a second scale parameter is necessary to fully describe the anisotropic strain conditions.

2.2 Strain analysis

Due to the small displacements, the six transformation parameters usually are not meaningful (i.e. a clear distinction of the several blocks is not possible out of these parameters). On the other hand the affine transformation is analogous to a strain analysis assuming homogeneous and infinitesimal strain (e.g. [5], [2]). So there is a direct relation between the transformation parameters and the infinitesimal strain components e_{xx}, e_{yy} (rate of change of length per unit length in direction of x-axis respectively y-axis), e_{xy} (= e_{yx}, rate of shear strain) and the derived rotation angle ω. A better interpretation of the strain parameters can be reached by the transformation into the principal strain axes system, represented by the strain ellipse (Tissot indicatrix). The elements of the strain ellipse (the semi-axes e_1, e_2 and the orientation θ of the maximum strain rate) are calculated from the strain components analogous to the geodetic point error ellipse.

In contrast to Welsch [5], who recommends not to use translation parameters, here an integration of the translation parameters (t_y, t_x) seems useful to fully describe the movement pattern of a block to avoid that the strain parameters are distorted due to translational movement.

3 Fuzzy System

For the implementation of the algorithm MATLAB was used. It provides an initial fuzzy system, which supports all necessary calculation methods

(membership functions, methods for fuzzification, inference system, defuzzification). The developer of the fuzzy system has to choose the suitable methods, input and output parameters for his application.

3.1 Input variables

The indicators used as input parameters in the fuzzy system can be determined by the results of the sequence of affine transformations. Some of the indicators are presented here:

- The standard deviation s_0 is used for a first evaluation of the block properties. Investigations have shown that especially the change of the standard deviation from one step to the next one is a very important indicator for the assessment if the block is still 'correct'.
- The next parameters for the fuzzy system are the two semi-axes of the strain ellipse (TISSOT indicatrix): e_1, e_2. In case a point does not belong to the block investigated, e_1 and e_2 become significantly large. Here, the absolute values on the one hand and the change of the parameters between two subsequent steps of iteration on the other hand are used.
- Investigations have shown that the variation of the residuals is a good indicator to distinguish between a correct block (all points belong to one block) and an incorrect block (point of a neighbouring block is included). The interquartile range (used in the exploratory data analysis) gives an estimate of the variation of the residuals in every step of the transformation (see figure 2). The interquartile range was analysed for 170 cases consisting of 4 points. The interquartile range of the 31 correct blocks (= 4:0) is significantly smaller than for the incorrect blocks (81, 58 cases respectively). '4:0' in figure 2 denotes that all 4 points are lying on one block, '3:1': 3 points on one block, 1 point on the neighbouring block, '2:2': 2 points per block.

3.2 Membership functions

To use the input values in the fuzzy system, the 'sharp' values have to be fuzzified. Therefore, membership functions are used. MATLAB provides many membership functions, such as piecewise linear functions, the Gaussian distribution function,...

For most of the input parameters the membership function type 'trapezoidal' was chosen in a rather intuitive way. The inputs are split in 3 membership functions (per input). An example for the input parameter 'change of standard deviation s_0' is given in figure 3. It is represented by three membership functions of type 'trapezoidal': small, medium and big.

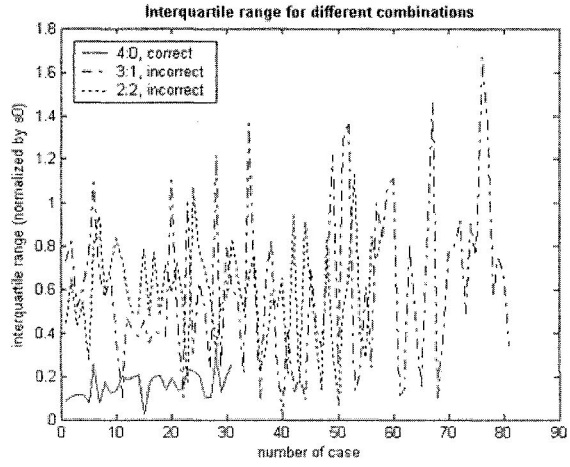


Fig. 2 Interquartile range for 170 different cases.

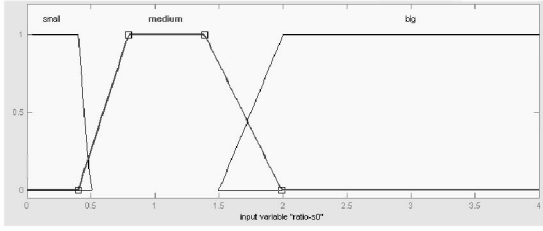


Fig. 3 An example for the membership functions of an input parameter.

3.3 Rules

The rules used in the inference system have been found empirically. At the moment, 25 simple rules have been implemented. But this topic is still under investigation. The rules implemented must have a structure like: If (input is X) then (output is Y), followed by a weight for this rule, e.g. (1). Examples for some rules used in the inference system are:

If (e_1 is small) then (termination is unlikely) (0.5)
If (e_1 is medium) then (termination is indifferent) (0.5)
If (e_1 is big) then (termination is likely) (0.5)

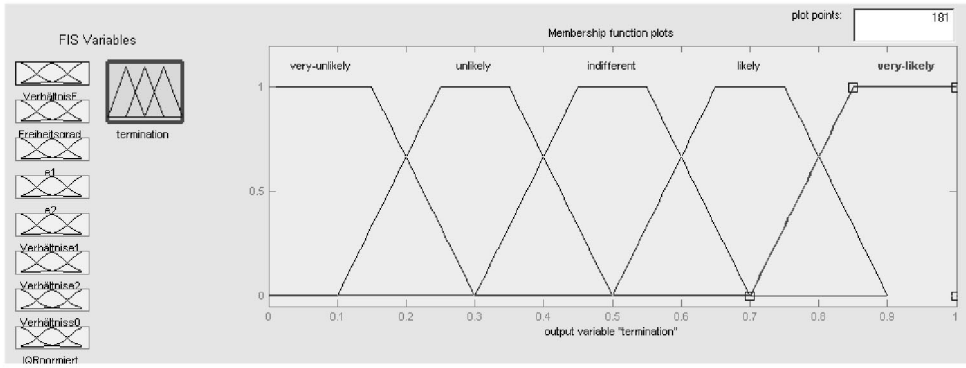


Fig. 4 The output value (probability for termination) is represented by 5 membership functions: very unlikely, unlikely, indifferent, likely, very likely.

3.4 Output

After the evaluation of the actual values of the input parameters, a fuzzy set for each output variable exists that needs defuzzification to get a single, 'sharp' output value.

Here only one output is implemented. It represents the probability that the block is complete, i.e. that no point in the neighbourhood fits the block under investigation so that the search algorithm has to be stopped. The output value is represented by 5 membership functions, see figure 4.

4 Example

In this section the algorithm is tested with the so-called Delft network. The simulated measurements (distances and directions) for different epochs (1, 2a, 2b, 3a, 3b) can be found in [6]. In this testing scenario, epochs 1 and 3b were used to calculate a geodetic deformation analysis using the software PANDA [4]. The displacement vectors, which are the results of the deformation analysis, are the input for the block detection algorithm (see figure 5 and table 1).

Table 1. Displacements dy and dx between epochs 1 and 3b.

Point	dy [m]	dx [m]	Point	dy [m]	dx [m]
3	0.232	0.158	35	-0.054	-0.121
5	0.197	0.180	37	-0.035	0.004
11	-0.021	0.339	39	0.113	0.248
13	-0.072	-0.069	41	0.042	0.270
15	-0.087	0.032	43	-0.018	-0.181
17	-0.058	0.033	45	0.002	-0.186
21	0.025	-0.166	47	0.001	-0.124

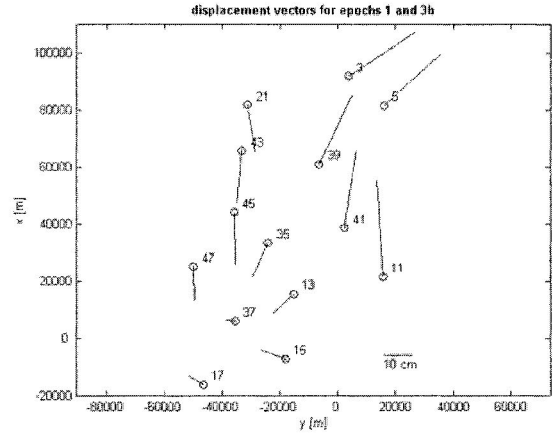


Fig. 5 Graphical representation of the displacement vectors for the epochs 1 and 3b of the Delft network.

The calculation of all blocks consisting of 4 neighbouring points gives 170 possible cases. To find the first minimal block, these cases are sorted by the standard deviation, under the condition that the values for e_1 and e_2 are within user-chosen limits (it is necessary to consider the actual material properties).

So the first block determined by the algorithm consists of the points 3, 5, 11, 41 with the standard deviation s_0 = 9.9 mm. In the next step all the neighbouring points are used to find the next point of the block:

Possible candidates	13	15	21	35	**39**	43
std dev s_0 [mm]	78.3	59.5	52.6	66.2	**12.7**	55.8

The algorithm decides that the best candidate for a fifth point in the block would be point 39 because of the minimal standard deviation s_0 = 12.7 mm. The fuzzy system now calculates all the necessary input parameters and after evaluation of the implemented rules it concludes with the following output (figure 6).

The defuzzified output value is 0.30, that means a probability of 30 % that the block should be terminated. So the next step of iteration is started to find the sixth point of the block.

Possible candidates	13	15	**21**	35	43	45
std dev s_0 [mm]	94.4	72.3	**71.3**	86.9	76.4	76.7

In this case, point 21 is the best candidate for the block (note the rate of change of s_0 between the two steps of iteration). But the fuzzy system calculates an output value of 0.72, which means that the block should be terminated with a probability of 72 %. So point 21 has to be removed from the selected block.

Now the first block is complete. The algorithm starts again the search for a minimum block of 4 points. The combination of points 13, 15, 17, 35 with a standard deviation s_0 = 12.0 mm is chosen. The results of the next steps can be found in table 2. After 6 iterations the algorithm stops because all points have been used.

Table 2. Results of the subsequent steps of the algorithm. Point 21 is the last point, so the algorithm is terminated (neglecting the output value of the fuzzy system).

Iteration number	Points included in the actual block	s_0 [mm]	Probability of termination
1	13 15 17 35	12.0	
2	13 15 17 35 47	11.9	0.30
3	13 15 17 35 47 45	11.0	0.30
4	13 15 17 35 47 45 37	17.1	0.39
5	13 15 17 35 47 45 37 43	24.3	0.48
6	13 15 17 35 47 45 37 43 21	31.6	0.50

As a result, two blocks have been identified. This result corresponds to the pattern that people would intuitively find after viewing the graph of the displacement vectors (figure 5 above).

The different block detection algorithms presented in [6] produce the same blocks as well.

5 Conclusion and Acknowledgement

In the past years, fuzzy methods were used more and more in geodesy (e.g. [7]). Fuzzy systems are a good method for the assessment of imprecise data or for the processing of linguistic variables because it is possible to reproduce the human way of decision making.

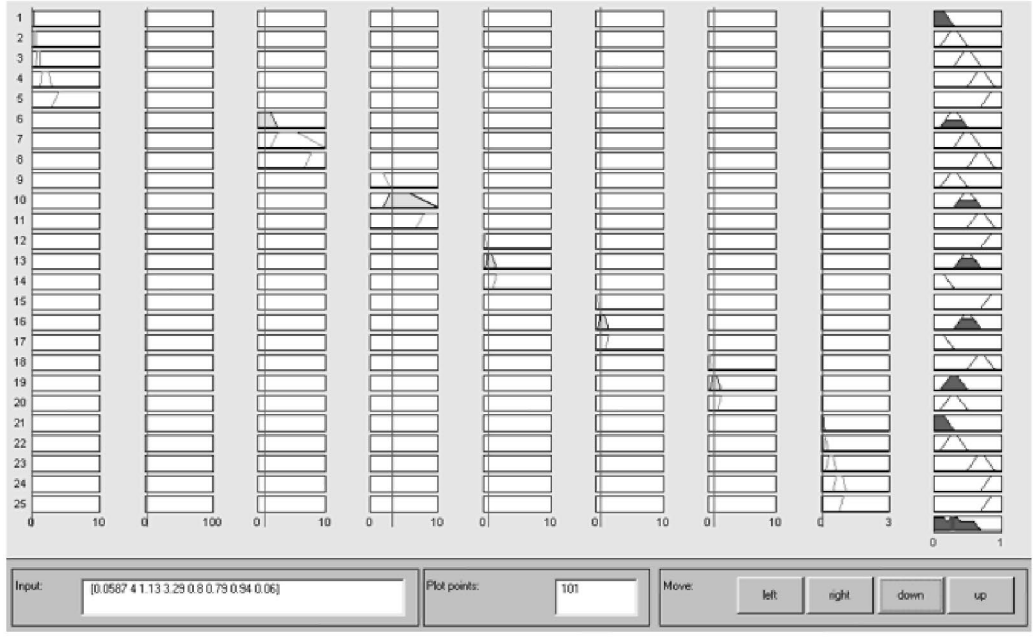


Fig. 6 Screenshot of the aggregation and defuzzyification process for the first run of the fuzzy system; output value: 0.30.

For the movement of a sliding area, it is hardly possible to build a suitable mathematical model fully describing all of the complex processes in the background (geology, hydrology,...). So this situation is well suited to be treated by fuzzy methods, where the knowledge of different sciences can be combined in a non-formal way. Here a possible assessment algorithm for landslide monitoring data is developed. The example of the Delft network shows that the problem given can be solved using a fuzzy system. But there is much more work to do; it is planned to include other input parameters like e.g. geological information (type of material of the area investigated), so the number of input parameters and rules will increase. Additionally, the fuzzy system will be tested on further examples.

This work was partially supported by the European Commission, Research DG, Environment Programme, Global Change and Natural Disasters.

References

[1] A. Antonopoulos, W. Niemeier: Formulierung und Test impliziter linearer Hypothesen bei der geodätischen Deformationsanalyse. In: W. Welsch (Hrsg.): Deformationsanalysen '83. Beiträge zum Geodätischen Seminar 22. April 1983. Schriftenreihe Wissenschaftlicher Studiengang Vermessungswesen Hochschule der Bundeswehr München, Heft 9. München, 1983.

[2] F. K. Brunner: On the analysis of geodetic networks for the determination of the incremental strain tensor. Survey Review Vol 25 , No. 192 (1979).

[3] H. Kahmen, W. Niemeier: OASYS: Integrated Optimisation of Landslide Alert Systems. Oesterreichische Zeitschrift fuer Vermessung & Geoinformation (VGI), 1/2003, ed. by H. Schuh, F. Brunner, and H. Kahmen, pp. 99 - 103, 2003.

[4] Panda: Handbuch zum Programm Panda (Programm zur Ausgleichung von geodätischen Netzen und zur Deformationsanalyse). GeoTec GmbH, Laatzen, 2002.

[5] W. Welsch: Description of homogeneous horizontal strains and some remarks to their analysis. In: Proceedings of the International Symposium on Geodetic Networks and Computations. DGK Reihe B, Heft Nr. 258/V. München, 1982.

[6] W. Welsch (Hrsg.): Deformationsanalysen '83. Beiträge zum Geodätischen Seminar 22. April 1983. Schriftenreihe Wissenschaftlicher Studiengang Vermessungswesen Hochschule der Bundeswehr München, Heft 9. München, 1983.

[7] A. Wieser: Robust and fuzzy techniques for parameter estimation and quality assessment in GPS. Dissertation, TU Graz. Shaker Verlag, Aachen, 2002.

[8] U.S. Geological Survey: http://landslides.usgs.gov/ html_files/landslides/program.html

OASYS - Integrated Optimization of Landslide Alert Systems

H. Kahmen
Institute of Geodesy and Geophysics, Department of Applied and Engineering Geodesy, Vienna University of Technology, Gusshausstrasse 27-29, 1040 Vienna, Austria

W. Niemeier
Institute of Geodesy and Photogrammetry, University of Technology Braunschweig, Gaußstr. 22, D-38106 Braunschweig, Germany

Abstract. In the past there has been a wide range of research work on landslides. However, most of this research has been carried out by a single discipline such as geology or geodesy. Therefore an international and interdisciplinary project was started. The proponents of the project, sponsored by the European Union, believe that a multidisciplinary integration of different methods has the greatest potential for substantial progress in natural hazards management. The goal of the project is the development of methods that allow: The detection of potential landslides, an efficient and continuous observation of critical areas, the derivation of real time information about actual risks.
This can be achieved by three steps:

- Detecting of potential landslides by large scale monitoring.
- Detecting of taking-off-domains by local scale monitoring.
- Impact and risk assessment and development for strategies for alert systems.

Keywords. Natural hazards, landslides, alert systems

1 Introduction

Currently, landslides are a major natural hazards around the world, killing or injuring a large number of individuals and creating very high costs every year.

Besides direct costs, landslides are also responsible for indirect costs like the interruption of important infrastructure facilities or tourist industry losses etc. In the future it is very likely that the damages caused by landslides will increase as the hilly areas, where the majority of the landslides occur, are used by a growing number of tourists, and intersected by increasingly powerful transnational networks. In addition many global climate change scenarios predict an increase in the probability of heavy rain, which is a primary trigger for landslides. This implies that there is urgent need for highly productive and reliable tools for landslide hazard management at an operational level.

In contrast to these prospects, the national funds for hazard prevention and the number of personnel in the national authorities responsible for hazard management are stagnating or shrinking. Several natural disasters in the last years have also shown that the decision makers in hazard management are confronted with a public that will evaluate their decisions in an increasingly critical manner, and look very thoroughly for possible responsibilities.

There has already been a wide range of research work on landslides [1] [2] [3]. Most of this work had a bias towards one discipline, such as geodesy or geology. The proponents of the project believe that for future work the multidisciplinary integration of different methods has the greatest potential for substantial progress in natural hazards management. The aim of the suggested research project is to set up an integrated workflow for landslide hazard management. This system should lead the practitioner from data acquisition to suggestions of risk management measures. The majority of fatalities occur due to the fact that the affected persons are not aware of the danger, and are surprised then by the actual landslide event.

The emphasis of the suggested project is the development of observation methods that allow:

- The detection of potential landslides.
- An efficient and continuous observation of critical areas.
- The derivation of real time information about actual risks.

2 Existing technology

Currently the investigation of landslides and unstable slopes is based on two groups of information sources, at a regional and local scale. The information sources on reginal scale are the following:

- **Data on historical landslides:**
  The relevant data should include information when and where landslides occurred in the critical areas.
- **General conditions of the areas:**
  Map data of the general conditions, like digital elevation model, geology, tectonics, geomorphology, vegetation, climate, land use are indicators for active landslides [4].
- **Remote sensing data:**
  Further sources for the potential landslide risk can be differential satellite image analysis, time series of airborne sensor data (photogrammetry, laser scanners, radar systems, geophysical measuring devices) and terrestrial mapping.

The information sources on local scale are the following:

- **Geodetic:** GPS, precise levelling, tacheometers (measurement robots), multisensoral remote sensing techniques (using optical- and radar-sensors)
- **Geophysical:** geoelectric and electromagnetic field measurements.
- **Geodynamic:** borehole tiltmeters, extensometers, hydrostatic tiltmeters.
- **Seismological:** sensors for microseismic activity, seismic reflexion measurements.
- **Hydrological:** sensors for groundwater level variations, water level variations, groundwater stream variations.
- **Meteorological:** sensors for temperature, air pressure, precipitation.

The information of the regional source enables the detection of the landslide site, those of the local source are used to describe the mechanism of the processes. Normally, however, only some of the parameters (deformation vectors, water power pressure) are used for the investigations and the measurement points are widely spaced, and divided across the landslide domain. The deformation measurements are frequently based only on one measurement method (for example GPS or tacheometry). Since these measurement methodologies are relatively costly, normally only a limited number of observation points are used and the measurement systems are not run continuously. This methodology has made it possible to follow the evolution of the landslides precisely, but it is difficult to predict the evolution.

An improved integration of information sources would allow for more reliable predictions in the future.

3 Integrated Optimization

An advanced model is now under investigation, based on large scale monitoring and a multi-component alert system. The method will consist of three different steps.

3.1 Step 1, Detection of potential landslides (Large scale monitoring)

To obtain information about long-term geodynamical processes, a large scale evaluation has to be performed. This includes historical data, all information describing the general conditions of the areas (as outlined in section 2) and remote sensing data, such as aerial photographs and optical and radar images from satellites (Fig. 1).

More specific techniques such as remote sensing (InSAR), differential GPS (using phase observations) and tacheometric measurements will provide additional information about the deformation process (Fig. 2). The measurements will be taken only three or four times a year, and the results will be fields of vectors describing the displacements and velocities [10].

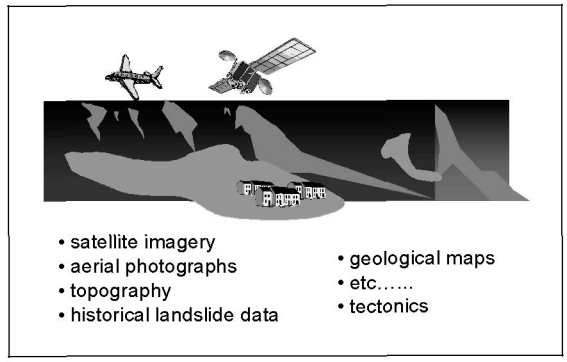
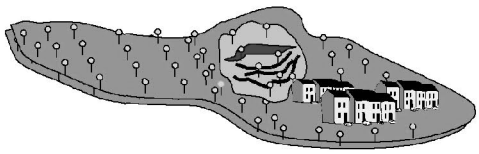


Fig. 1. Large scale monitoring with airborne and satellite remote sensing.

An advanced and generalised deformation analysis algorithm based on geometrical, topographical, geological, hydrological and meteorological information has to be developed to

improve the detection of taking-off domains, see [7].



Coarse observation network for identification of critical areas

Fig. 2. Large scale monitoring using GPS and total stations.

3.2 Step 2, High precision relative measurements

High precision borehole tiltmeters, extensometers, hydrostatic levelling and further relative measurement systems shall be used in the area of the taking-off domains to get online information about the geodynamical process [6] [8].

This multi-sensor system will be running continuously and can therefore support a real time alert system, see Fig. 3 [5].

3.3 Step 3, Impact and risk assessment/ Development of strategies for alert systems

Risk assessment comprises the analysis of the empirical data, and the development of an alert system deals with the management of an impending landslide hazard.

The analysis of empirical data includes the detection and interpretation of velocity fields in order to define zones of increased deformation. The final analysis will be supported by expert systems, using methods of cluster analysis, neuronal networks, fuzzy techniques and others [7].

The process of risk analysis can be divided into hazard analysis and vulnerability analysis. Hazard analysis is the review of the potential hazardous processes. Scenarios for the evolution of a landslide area of interest have to be set up, including an estimation of the probability of these scenarios. Different scenarios will affect different areas and therefore a different amount of people and property. The assessment of the impact of different hazard processes on the affected population and its property is called vulnerability analysis.

The integration of hazard and vulnerability analysis will lead to an estimation of the actual risk situation of the affected population. The risk management measures will depend heavily on the specific conditions and will include landuse planning, technical measures (e.g. drainage systems), biological measures (e.g. afforestations) and temporary measures (e.g. evacuations).

Informing of the public about potential hazards, and the fast and effective alerting and evacuation of the people in endangered areas are vital prerequisites for a functioning natural hazard management system. One of the main sources for fatalities caused by natural hazards, is the lack of awareness of certain hazards and the lack of knowledge about appropriate behavior in an emergency situation. This work package tackles this deficit by the dissemination of background information to the public, and by the development of new means of alerting and evacuation by the application of new communication technologies. This has to be performed by taking into account the cultural and socioeconomic boundary conditions in each specific case. Participative models for hazard mitigation/prevention measures will be developed [9].

An advanced alert system will consist of the following components:

- Communication system which transforms the information to a control station.
- Communication system which informs the persons who have to make decisions.
- Quality criteria to support the decisions.
- Communication system which informs the persons who have to start actions.

Information transfer to the affected population

Fig. 3. Continuous local scale measurements and the information transfer.

4 The integrated work flow

The integrated workflow for landslide hazard management, as outlined in section 3, is depicted in

Fig 4. The activities in the regional scale comprise the evaluation of historical landslide data, data of different geo-sciences and meteorology, and monitoring data. As a result a multidisciplinary detection and calssification of landslides shall be possibel. This will finally provide sufficient information for risk assessment, a multiscale optimization of monitoring concepts for alert systems, and for land use plans. The easy to use alert system shall be based on a knowledge based system, so that numerical and descriptive data can be used. The integration of the acquired data in a multidisciplinary approach in order to detect and classify landslides will be one of the most challenging tasks of the project.

5 Final remarks

The project OASYS is supported by the European Commission under the 5th Framework in the area "Energy, Environment and Sustainable Development Programme" under contract EVGI-2001-00061. The project started January 1, 2003. A multidisciplinary working group was formed:

- Vienna Consulting Engineers (VCE), Austria
- Institut for Geodesy and Photogrammetry, Technical University Braunschweig, Germany
- Institut for Geodesy and Geophysics, Technical University Vienna, Austria
- Geodetical and Geophysical Research Institute of the Hungarian Academy of Sciences, Hungary
- Bureau for Applied Scientific Remote Sensing, Germany
- Geodata, Engineering Company, Austria
- Bureau of Investigations and Survey, China
- Hungarian Ministry of Interior, National Directorate General for Desaster Management, Hungary
- Egnatia Odos S.A., ConstructionCompany, Greece
- University of Modena, Italy
- School of Geodesy and Geomatics, Wuhan University, China
- Laboratory of Geodesy, Uni versity of Thessaloniki, Greece

With this group a network was established by which comprehensive research on landslides can be carried out during and after the project.

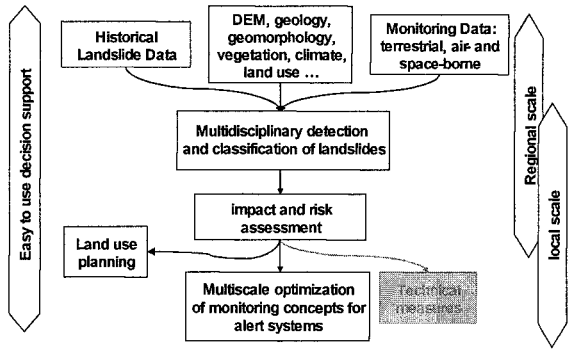


Fig. 4. Integrated work flow for integrates landslide hazard management.

References

[1] Almeida-Teixeira M.E., et al.(Eds.), 1991: Natural hazards and engineering geology. Prevention and control of landslides and other mass movements. *Proceedings of the European School of Climatology and Natural Hazards course* held in Lisbon from 28 March to 5 April 1990

[2] Bhandari R. K., 1984: Simple and economical intrumentation and warning systems for landslides and other mass movements. State-of-the-art lecture, *Proc. 4th Int. Symp. on Landslides*, Toronto, vol. 1, 251-273.

[3] Carrara A., 1984: Landslide hazard mapping: aims and methods, Mouvements de terrain. Association Francaise de Geographie Physique. Colloque de Caen, p. 141-151.

[4] Cui Zhengquan, 1999: Geology and Geomechanics of the Three Gorges Projects. Bureau of Investigations & Survey, Changjiang Water Resources Commission, MWR, May 30 1999

[5] Einstein H. H., 1988: Landslide risk assessment procedure. Special lecture in *Proceedings of the fifth international symposium on landslides*, Lausanne, p. 1075-1090.

[6] "Geotechnical Instrumentation for Monitoring Field Measurements", *Conference at University of Florida*, March 2001

[7] Haberler M., 2003 A fuzzy system for the assessment of landslide monitoring data, Österreichische Zeitschrift für Vermessung & Geoinformation (VGI), 1/2003, ed. by H. Schuh, F. Brunner, and H. Kahmen, pp. 92 - 98.

[8] Savvaidis P., Lakakis K., Zeimpekis A., 2001: Monitoring Ground Displacements at a National Highway Project: The Case of "Egnatia Odos" in Greece. *Proceedings of IAG Workshop "Monitoring of Constructions and Local Geodynamic Process"*, Wuhan University, China, 2001

[9] Spiker, E. C., Gori P. L., 2000: National Landslide Hazards Mitigation Strategy. *Open-file Report 00-450, USGS.*

[10] Theilen-Willige, B., 1998 : Seismic Hazard Localization based on Lineament Analysis of ERS- and SIR-C- Radar-Data of the Lake Constance Area and on Field Check.- *European Conference on Synthetic Aperture Radar*, 25- 27 May 1998 in Friedrichshafen, Proceedings, 447-550, VDE-Verlag, Berlin.

Combined Evaluation of Geodetic and Geotechnical Data during Tunnel Excavation by Use of a Knowledge-Based System

K. Chmelina
Geodata ZT GmbH, Subsidiary Vienna, Geyschlägergasse 14, 1150 Vienna, Austria

H. Kahmen
Institute of Geodesy and Geophysics, Department of Applied and Engineering Geodesy, Vienna University of Technology, Gusshausstrasse 27-29, 1040 Vienna, Austria

Abstract. The paper presents three examples of how geotechnical knowledge can be used for the analysis of geodetic deformation data. The knowledge is represented as facts and heuristic rules which are implemented in a knowledge-based system. The system is supposed to support the geotechnical expert on site by an automatic detection of specific deformation phenomenas. With this contribution the application field of NATM-tunnelling (New Austrian Tunnelling Method) is addressed.

Keywords. Knowledge-based system, deformation analysis

1 Introduction

The geodetic monitoring of 3D displacements during tunnel excavation has become a standard procedure for tunnel projects. In NATM-projects (New Austrian Tunnelling Method [3]), the immediate geotechnical interpretation of the periodically observed displacements must be seen as an integral part of the tunnelling method. As a consequence, qualified and experienced geotechnical experts are needed on site.

Currently, geotechnical interpretation is based mainly on the daily analysis of numerous types of sophisticated displacement diagrams. Together with many other listings and graphics showing further project data, this is a time consuming manual process. For up to date analysis of large data volumes or online analysis of automatically monitored data, e.g. when robotic tacheometer systems are used, this manual procedure is obviously impracticable. Until now, there exists no means by which these tasks are managed quickly, automatically and with sufficient reliability.

A knowledge-based approach is presented which will assist interpretation work through automatic analysis of the displacements by use of heuristic rules. These rules provide expert knowledge for the detection and interpretation of specific deformation phenomenas, and thus building the core of a knowledge-based system [2]. Three different examples using geotechnical knowledge to evaluate geodetic data, are presented.

2 Example ONE: How to Find an Explanation for Reactivated Deformations

In general, any *unexplainable* deformation behaviour leads to an uncertainty, which must be solved quickly (the causes of the deformations must be found) in order to ensure safe continuation of the project. Such a case can occur when deformations that have been consolidated in the past suddenly become reactivated. Whether such a reactivation is really *unexplainable* depends on the current situation on site. In the case of NATM-tunnelling, deformations must be analysed together with the excavation progress. If, for example, an excavation face (e.g. bench, invert) approaches or passes the deformation area (situated in or above a tunnel) the reactivation of deformations usually is expected and explainable. On the other hand, any reactivation that cannot be explained (e.g. by excavation activities) is alarming and perhaps critical for the construction progress. Fig. 1 characterizes the geotechnical interpretation steps, which handle this problem.

Modelling these steps by conventional programming would lead to a complex code consisting of numerous interlocked IF – THEN statements, logical expressions (AND, OR, NOT) and loop commands. Therefore a rule-based system was chosen here. With this system the interpretation process can

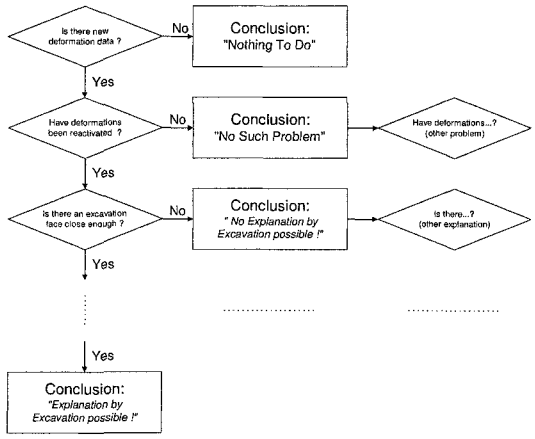


Fig. 1 Interpretation steps

be modelled in a modular, transparent and more flexible way. With respect to a better understanding of rule-based inferencing the use of WHENEVER instead of IF is more appropriate. The decisive difference and advantage is that by use of rules the interpretation must not necessarily be processed top-down. The developed rule-based system can ask all the questions shown in Fig. 1 dynamically at the same time but does not have to ask them sequentially (as with IF-THEN). Further, the questions can remain permanently active in memory but do not have to be repeatedly asked (as with loops). To elucidate let's transfer the problem into heuristic rules:

Rule 1:

WHENEVER	deformation data match the pattern 'Reactivation'.
THEN	produce a fact that marks this data as reactivated.

The precondition of the rule consists of a representation of an abstract deformation pattern called 'Reactivation'. At implementation level this pattern is defined in a more or less empirical way as a time series of deformation data, showing increasing deformation velocities during the last 3, 4 or 5 epochs. The action of the rule is to add (assert) meaningful information (a fact) to the working memory describing which point, which epoch, etc. is concerned. This is completed for all such incidents, thus the knowledge-based system becomes aware and mindful of them.

Rule 2:

WHENEVER	1. there is a fact that marks data as reactivated. 2. excavation data match the pattern 'Close Enough'.
THEN	produce a fact that describes this link.

The rule consists of two preconditions to be regarded as connected by the logical AND-statement. The first is fulfilled if there exists at least one of the previously mentioned facts produced by rule 1. The second consists of another abstract pattern 'Close Enough' specifying that excavation data prove excavation activities to be a possible cause for a detected reactivation incident. For that, certain distances in space and time constraints must be considered, which are defined according to assumptions of the expert. If both conditions are fulfilled the rule adds a meaningful fact to the working memory, expressing that there is (possibly) a connection between reactivation and excavation. The rule fires for each excavation activity meeting 'Close Enough', and links it to the reactivation incident (e.g. a reactivation can be explained by more than one excavation activity). As can be seen, linking is performed whenever new excavation data or new deformation data is accessible. So rule 1 and rule 2 together keep track of both data.

For continuing interpretation (Fig. 1) rule 3 could be specified by simply modifying the second precondition of rule 2, but now assuming the opposite case that there is no excavation data matching the pattern 'Close Enough'. Thus the corresponding 'No Explanation' – conclusion could be processed and stored. This would mean that the normally large volume of excavation data has to be checked twice for each reactivation incident. To speed up runtime performance, rule 3 is alternatively designed as follows:

Rule 3:

WHENEVER	1. there is a fact that marks data as reactivated. 2. there is no fact describing a link to excavation data.
THEN	conclude (display and store) that there is no explanation.

Adopting this method, the system must know that it cannot begin applying rule 3 before rule 2 has been used, which serves as an example that asking all questions at the same time is not always efficient. This meta-knowledge was implemented by an ap-

propriate hierarchical structuring of rules. The majority of rules were kept at equal level.

By additional rules the knowledge-based system is able to continue forward reasoning down to a theoretical final conclusion:

'The phenomena could have been caused by the excavation',

which may or may not be reached. At the same time further interpretational problems can be handled by other rules, allowing for similar *horizontal* and *vertical* conclusions. If the user is not really interested in all problems he/she can preselect the desired ones in a configuration dialog. The selection is then transferred into facts triggering only the relevant heuristic rules.

The system can be called heuristic as the arrangement and composition of the rules was mainly driven by common sense, efficiency aspects and some good advices from experts. System tests proved that the software can process even large data volumes within acceptable times. What has to be observed more critically is the error budget. Wrong conclusions might be drawn, especially when there are inaccuracies, errors and/or missing data. This problem was realized at an early stage, and as a consequence, a knowledge-based error detection method also based on heuristic rules was implemented (see example 2 below).

With regard to the problem of wrong conclusions, it is emphasized that the original intention was to develop a support system where the human expert still has the last comment. All final conclusions of the system are therefore reported in the form of insecure statements ('could').

3 Example TWO: How to Find and Analyse an Error in Deformation Data

As mentioned, the knowledge-based system must reason over actual measuring results. In order to avoid wrong conclusions due to bad quality of input data, error detection is needed in advance. (In practice the geotechnical expert does not consider error detection as a separate interpretational task - in other words: It's not their job to find errors.) So as long as data quality problems are seldom he/she apriori trusts in the data provided. He/she is however able to recognize (possibly) incorrect data by specific characteristics or signs mostly driven by experience and common sense. Further, to be sure, he/she begins to investigate the problem in detail. A typical error investigation strategy is briefly described as follows:

1. Discovering a characteristic error sign (e.g. an unexpected deformation) more or less by accident.
2. Determining its relevance for the interpretation.
   If the sign is rather small it might be considered irrelevant and neglected by the expert. If the sign is strong it still might be irrelevant (e.g. if the corresponding data has no influence on the interpretation at all).
3. Deriving a hypothesis (of a possible error cause, error type) from a known correlation between the sign and a cause.
   If various causes could be responsible the *best* one is chosen first (the most obvious, frequent or probable, the easiest to be checked, etc.). Now the sign becomes a symptom.
4. Searching for further symptoms supporting the hypothesis.
   The search is driven by the distinct intention to prove the hypothesis.
5. Searching for signs, facts, hints, features, etc. contradicting the hypothesis.
   Now the search is driven by the opposite intention to disprove the hypothesis. Which kind of search is started first depends on efficiency criterias. Investigations show that for psychological reasons experts usually try to verify (especially their own) assumptions first, even if their falsification would mean less effort.
6. Evaluating of all gathered aspects for deriving a verification certainty.
   The question is: 'Can the hypothesis be established as diagnosis ?' We may decide to stop and establish the hypothesis as diagnosis or to go back to point 3 deriving and evaluating an alternative hypothesis. If none of these can finally be verified sufficiently (better: satisfactory) either the initial error assumption is cancelled or the investigation result remains a list of possible but uncertain causes.

A knowledge-based system for handling this complex strategy needs an appropriate way to represent and propagate uncertainty – a problem that leads to the various concepts of probabilistic reasoning. Among those, the Certainty Factor Method [4] [2] was chosen because it represents a pure heuristic approach especially developed for rule-based expert systems.

Certainty Factors - a brief introduction:

Each premise (condition) $P_1, P_2,\ldots,P_n$ of a rule (e.g. the statement: 'There is an error sign') is attributed with a numeric parameter (a certainty factor) cf_1, $cf_2,\ldots,cf_n$ describing its degree of membership to the

concept *true*. The parameter interval lasts from –1 (not true) to +1 (true). A certainty factor $cf_x = 0$ means that there is nothing known about the truth of P_x. Also, each rule R_1, R_2,...,R_n itself (e.g. the statement: 'From the error sign can be concluded an error cause') has a certainty factor cfr_1, cfr_2,...,cfr_n expressing the strength (the truth) of the relation between the premises and the conclusion. When a rule R_x is executed, all involved certainty factors cf_1, cf_2,...,cf_n of its *n* premises and cfr_x are processed to obtain a certainty factor cfc_x for the conclusion C_x. In rule based expert systems a common way is to calculate:

$$cfc_x = min\ (cf_1, cf_2, \ldots, cf_n) * cfr_x . \qquad (1)$$

If by two different rules R_1, R_2 the same conclusion is drawn, the now concurring different values of cfc_1 and cfc_2 must consequently be processed to a final *cfc*, which can be completed as follows:

$$cfc = \begin{cases} cfc_1 + cfc_2 - cfc_1 * cfc_2 & if: cfc_1, cfc_2 \geq 0 \\ cfc_1 + cfc_2 + cfc_1 * cfc_2 & if: cfc_1, cfc_2 \leq 0 \\ \dfrac{cfc_1 + cfc_2}{1 - \min(|cfc_1|, |cfc_2|)} & if: cfc_1 * cfc_2 < 0 \end{cases} \qquad (2)$$

If further rules lead to the same conclusion the algorithm is applied sequentially. The mathematical operations for processing the certainty factors are thus chosen intuitively. They heuristically approximate the human way of propagating uncertainty.

An example is now given to imitate the above interpretation strategy by applying the Certainty Factor Method. It shall be analysed if there is a systematic height error of the instrument station in the measured vertical deformations. At first, rule 4 is designed for detecting an abstract pattern 'Strong Settlement'. Its conclusion (fact) is considered to be an initial and relevant error sign for a specific cause (steps 1, 2 and 3 of the strategy).

Rule 4:

WHENEVER 1. deformation data match the pattern 'Strong Settlement'.
2. there is no fact so far that marks the related epoch and station as suspicious.

THEN produce a fact that marks the related epoch and station as suspicious.

The rule fires exactly one time for all measuring epochs and instrument stations having at least one measurement showing the pattern. The system now knows all suspicious epochs and stations that have to be further investigated. Now (and only for them) there is applied the Certainty Factor Method. The following initial certainty factors for the following three statements (rules) are defined as:

$$cfr_1 = 0.3\ e^{-(bew_diff)^2} \qquad (3)$$

If there is a strong settlement incident the certainty factor of the error is cfr_1.

If there is only one measurement available (which shows the error sign) the error is concluded with a certainty of 0.3, thus staying rather uncertain. If there are more measurements showing the error sign, the certainty of concluding the error from one such sign becomes a function of the similarity of all signs. The function models the expectation that a height error of the station should be visible in all measurements in the same size. The less that one sign meets this expectation, the less it is a hint for this specific error. This aspect is implemented by multiplying 0.3 with an exponential term considering the difference between the size of the sign and the mean size of all signs. The rule models step 4 of the strategy producing a list of positive certainty factors.

$$cfr_2 = -0.1 \qquad (4)$$

If there is an inconspicuous measurement the certainty factor of the error is cfr_2.

If a measurement from the suspicious station appears as normal it is treated as if it would slightly contradict the hypothesis with a negative certainty of – 0.1.

$$cfr_3 = -0.5 \qquad (5)$$

If there is a measurement showing a strong uplift the certainty factor of the error is cfr_3.

If a measurement shows a strong uplift, which indicates the opposite than initially expected direction of deformation, it contradicts the hypothesis with a high negative certainty of – 0.5.

The last two rules model step 5 of the strategy, producing a list of negative certainty factors.

Finally, all certainty factors were processed by (2) giving a final certainty of the hypothesis. Fig. 2 shows an example of the vertical deformations of 7 points measured on 27 October for which the system states: 'A systematic height error of the station could have happened. For this kind of error a certainty of 0.87 was derived.' By studying Fig. 2 the reader is invited to compare this result with his/her own assumptions.

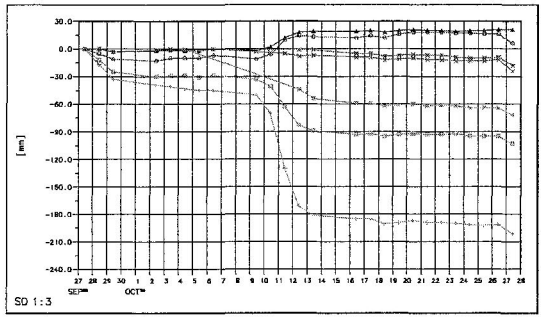


Fig. 2 Measured vertical deformations of seven points

By applying this method, the knowledge-based system is now able to offer an additional explanation for the interpretational problem discussed in section 2 (Fig. 1) which says:

'The phenomena could be caused by an error.'

4 Example THREE: How to find Remarkable Deformations by use of Rating Knowledge

During tunnel design, deformation prognosis is determined by use of diverse techniques (e.g. numerical simulations) providing the knowledge of how the planned tunnel (the rock) is expected to deform under certain conditions. Besides these *official* prognoses, the expert also has his/her own individual and much more extended expectations (usually derived from experience). During tunnel excavation the measured deformations are compared with the prognosed and expected ones. Unexplainable big differences between them can be an alarming sign for the geotechnical expert on site. An automatic detection of such a situation makes necessary the representation of the prognoses and expectations as well as the expert knowledge of how to rate their deviation(s) from reality. In the developed knowledge-based system, prognoses and expectations can be specified by defining deformation curves, vectors, values, etc. and assigning them to distinct measuring objects (points, sections, areas). The rating knowledge is represented by empirical rating models (Fig. 3), which can be selected, configured and assigned to a prognosis or expectation.

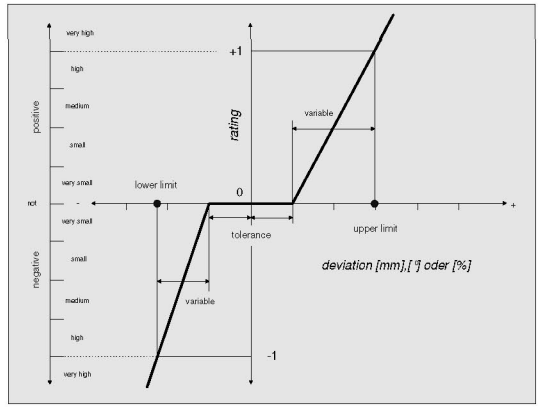


Fig. 3 Empirical Rating Model

The rating model allows us to transfer the difference between a real and prognosed (expected) measuring result (e.g. the difference in settlement [mm], the difference in vector orientation [°], the difference in deformation velocity [mm/d], etc.) into a numeric rating value, and an associated linguistic variable. The transfer function might either be linear (as exemplified in Fig. 3), exponential, logarithmic or determined by any other mathematic equation that best serves to approximate the users way of rating. In addition, a rating tolerance can be specified for maintaining the rating value equal to zero in case of small differences and an empirical measuring accuracy for putting the rating value to a worst-case value by simply adding the measuring accuracy to the difference before rating.

As soon as we input the knowledge into the system, heuristic rules apply it to the incoming deformation data. The rules ensure that all deviations exceeding user-definable evidence limits (e.g. all deviations reaching the *very high* - level) are detected and reported automatically. Rule 5 below shows the principle structure:

Rule 5:

WHENEVER	1. there is deformation data. 2. there is a prognosis for this data. 3. there is a rating model for this prognosis. 4. the calculated rating value exceeds an evidence limit.
THEN	produce a fact.

As can be seen, more than one prognosis may exist (e.g. from different numerical simulations) as well as more than one rating model (e.g. from different experts). However, the rule evaluates all of them, thus becoming aware of all evident cases. Alternatively, by a simple modification of the rule the knowledge-based system is able to offer another answer to the interpretational problem discussed in section 2 which says:

'The phenomena was predicted, maybe there is no problem at all.'

In order to provide more information about this kind of answer the system additionally outputs context information to the prognosis (prognosis by whom, when, based on what, etc.), which has to be input in advance (together with the prognosis).

5 Conclusion

It is shown how interpretation of deformations, occurring during the construction of a tunnel, can be based on a knowledge-based system. The system is thought to support the geotechnical expert on site by reducing his/her interpretation time and effort, which is necessary in the case of an automatic continuous deformation monitoring system. Classical deformation analysis methods, only based on geometrical data, cannot be used here, as in addition, information such as geological data, data about the excavation progress, data from simulation calculations, etc. have to be included in the evaluation process. This paper shows with three examples how heuristic rules can be developed and used in such a knowledge-based system. With the developed rules, deformations reactivated by the construction process, measurement errors and differences between predicted and measured deformation data can be analysed.

References

[1] Bonisone P, Decker K (1985). Selecting uncertainty calculi and granularity: an experiment in trading-off precision and complexity. In: *Proc. of Workshop on Uncertainty and Probability in Artificial Intelligence*, pp 57-66, Los Angeles, CA.

[2] Chmelina K (2002). Wissensbasierte Analyse von Verschiebungsdaten im Tunnelbau, Dissertation, Inst. of Geodesy and Geophysics, TU-Vienna.

[3] Müller L Fecker E (1978). Grundgedanken und Grundsätze der ‚Neuen Österreichischen Tunnelbauweise'. Clausthal, Trans Tech Publications.

[4] Shapiro SC (1992). Encyclopedia of Artificial Intelligence. John Wiley & Sons, Inc. Publishers - New York, 2nd edition.

SYMPOSIUM G02

Advanced Space Technology

C.K. Shum
P. Willis

Quasi real-time positioning of spacecrafts using the Internet VLBI system

T. Kondo, Y. Koyama, R. Ichikawa, M. Sekido, and H. Osaki
Kashima Space Research Center, Communications Research Laboratory,
893-1 Hirai, Kashima, Ibaraki, 314-8501, Japan;
e-mail: kondo@crl.go.jp; Tele: +81-288-84-7137

Abstract. We started the development of a quasi real time technique for spacecraft positioning by using the Internet-VLBI system in collaboration with the Institute of Space and Astronautical Science (ISAS). A series of VLBI observations receiving the signals from Japanese spacecraft GEOTAIL and NOZOMI have been carried out since June, 2002 to establish an observation method and to evaluate the measurement accuracy. At present it is possible to obtain the time delay which is a VLBI observable to use for satellite positioning within a few days after observations. Orbital determination software is under development in parallel to the improvement of observation data processing software. Phase delay measurement software is also under development to increase the measurement accuracy.

Keywords: Internet VLBI, Delta VLBI

1. Introduction

Communications Research Laboratory (CRL) has promoted a research on the quasi real-time spacecraft positioning by adopting a delta VLBI technique with the use of the Internet (Fig.1). A general-purpose VLBI system using a PC for data acquisition and the Internet protocol (IP) for data transmission has been developed as a part of this project. The system is called Internet VLBI or IP-VLBI system (Kondo et al., 2002).

The IP-VLBI system aims at a fully real-time VLBI operation from data acquisition to data reduction. At present time, automated observations and off-line correlation processing are available. A number of VLBI sessions observing telmetry signals from Japanese spacecrafts, GEOTAIL (Nishida, 1994) or NOZOMI (Yamamoto and Tsuruda, 1998), have been carried out using the IP-VLBI system since June, 2002 in collaboration with the Institute of Space and Astronautical Science of Japan (ISAS) and the National Astronomical Observatory of Japan (NAO) to check automatic observation software and to evaluate the accuracy of observed delay.

This paper presents the design of IP-VLBI system, fringes observed for various kinds of telemetry mode, and observed delay accuracy.

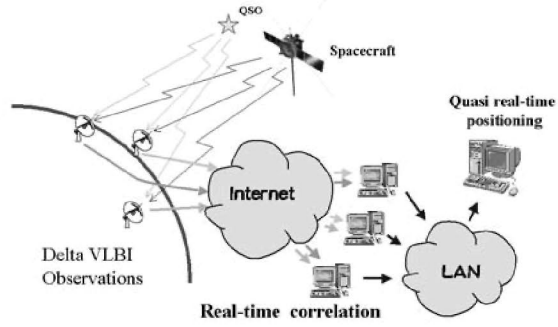


Fig.1. Concept of quasi real-time positioning of spacecraft using the Internet VLBI system.

Table 1. Specifications of a PCI VSSP board.

Reference signals	10MHz(+10dBm), 1PPS (TTL)
Number of input channels	max 4 ch
A/D resolutions	1, 2, 4, 8 bits
Sampling frequencies	40kHz, 100KHz, 200kHz, 1MHz, 2MHz, 4MHz, 8MHz, 16MHz

2. Internet VLBI system (IP-VLBI)

A new real-time VLBI system using Internet protocol (IP) technology has been developed to reduce network-cost and to expand connecting VLBI sites compared with an asynchronous transfer mode (ATM) based on a real-time VLBI system developed by the Keystone VLBI project (Koyama et al., 1998). We call this system "Internet VLBI" or "IP-VLBI".

The IP-VLBI system consists of a PC equipped with a PCI-bus Versatile Scientific Sampling Processor (VSSP) board. Table 1 summarizes specifications of VSSP board. The board has two reference signal inputs (10 MHz and 1 PPS signals) and four analog data input channels. Sampling frequency is from 40 kHz up to 16 MHz and A/D conversion resolution is from 1 bit to 8 bits. They are all programmable but maximum data rate is limited by the combination of CPU motherboard and PCI bus speed. So far we have attained the throughput of 64 Mbps stably. It corresponds to 4ch X 1bit X 16MHz sampling or 4ch X 4bit X 4MHz sampling, and so on.

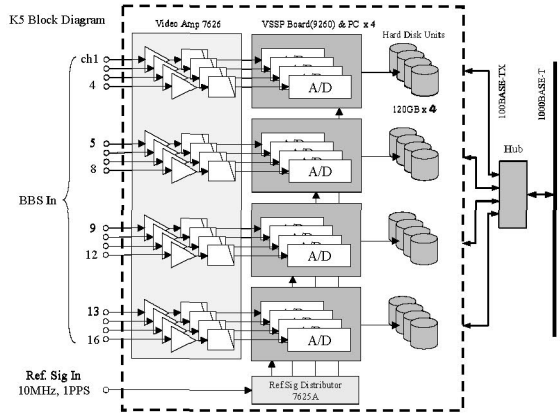


Fig.2. A schematic block diagram of K5-VSSP system.

We are expecting the increase of the maximum data rate optimistically, because the performance of PC and peripherals evolve quickly these days. A 64 bit-header data including precise time tag is inserted into the digital data stream every second. Sampled data are stored on a hard disk drive (HDD) and/or transmitted to a remote PC where software correlation processing is carried out. Correlated data are further processed to get multi-band group delay etc. A driver program for the VSSP board was first developed for FreeBSD system, but now it is available for Linux and Windows 2000 system.

The IP-VLBI system aims at taking over current geodetic VLBI system that receives 14 or 16 frequency channels at S and X bands. As a VSSP board on a PC can input 4 channel data, 4 PCs can cover channels necessary for a geodetic VLBI

Fig.3. A K5-VSSP prototype.

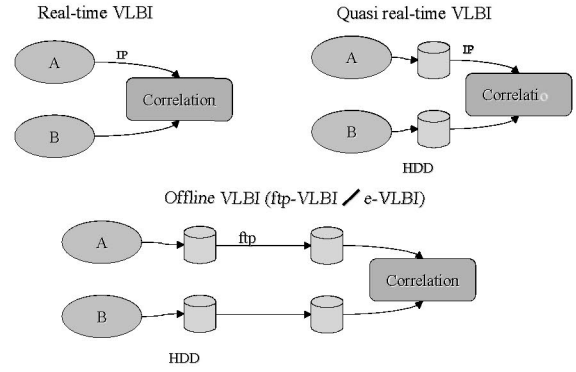


Fig.4. Supposed Operational Mode of the Internet VLBI

observation. The system assembled this way is named "K5-VSSP". A block diagram of K5-VSSP is shown in Fig.2. Each PC has four 120 GB HDDs for data storage, then total storage capacity is 1.92 TB. This corresponds to a 60000 sec scan length with 256 Mbps observation and is usually enough size to store a 24-hour (86400 sec) geodetic VLBI session, because antenna slewing time to switch radio sources is excluded. A picture of prototype K5-VSSP is shown in Fig.3.

Supposed operational mode of the Internet VLBI (K5) system is summarized in Fig.4. Although our final goal is real-time operation, off-line operation (FTP-based operation) and/or quasi real-time operation are the first requirement of system development. Regarding off-line operation, full automated observation is now available with K5.

Further system development is continued to realize a real-time VLBI operation.

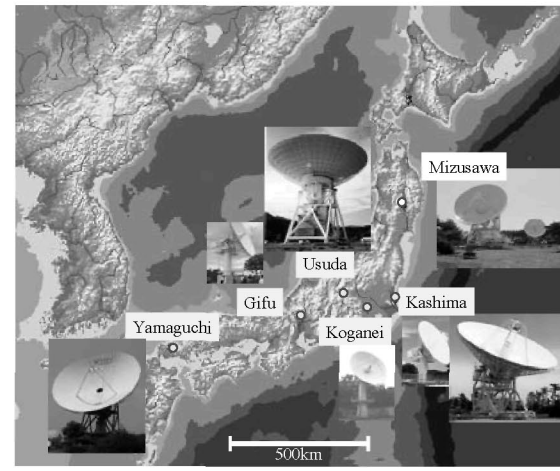


Fig.5 The location of stations participated in VLBI observations for satellite positioning.

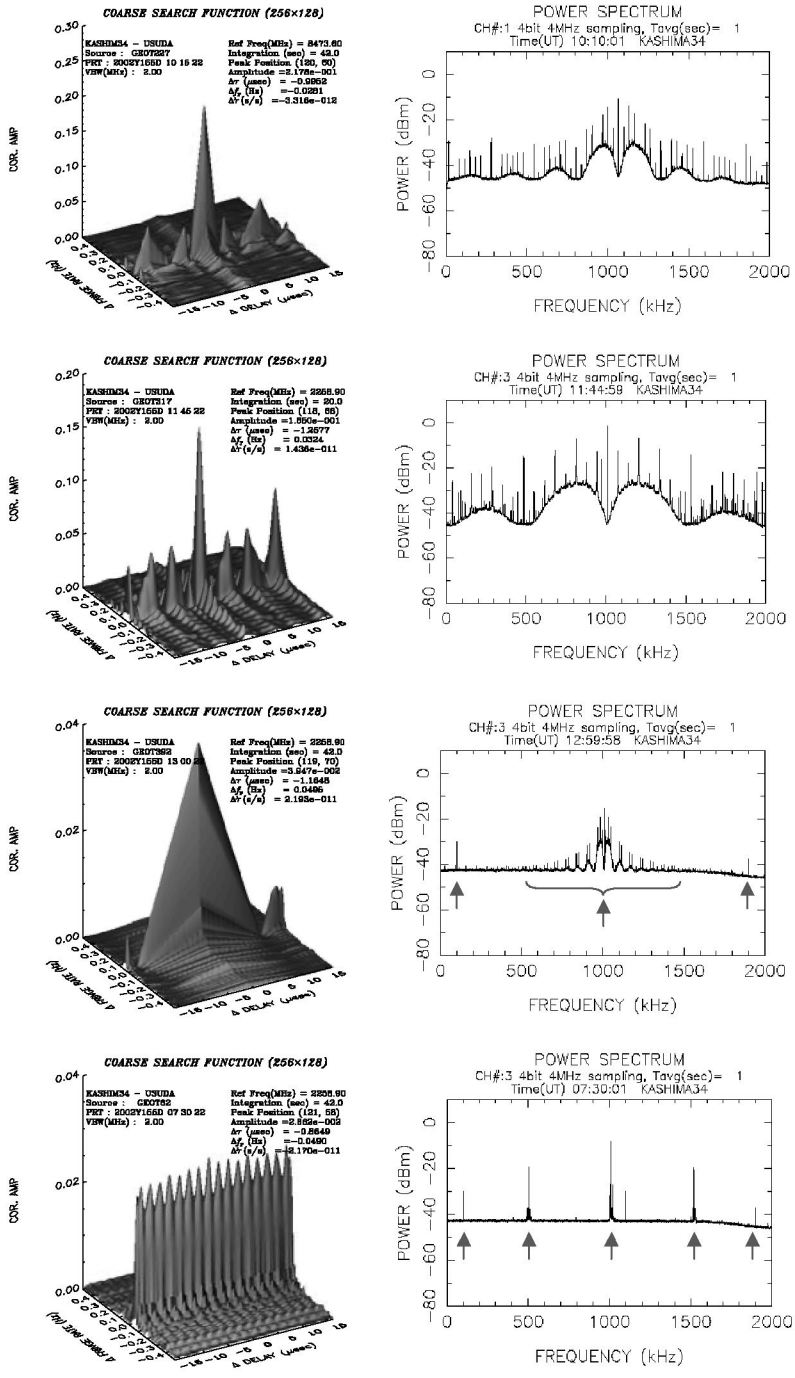


Fig.6. Fringes detected for GEOTAIL. Group delays can be determined without ambiguity excepting the lowest panel case.

3. Delta VLBI observations

The difference in the signal time of arrival at the two stations, referred to as the delay, is a VLBI's measure observable. This delay is composed of a geometric delay and other delays, such as, propagation medium delay, instrumental delays, clock offsets, etc. If a position well-known source is located nearby a target (spacecraft), differencing both observations will reduce common delays and only differential geometric delay applicable to a spacecraft position measurement

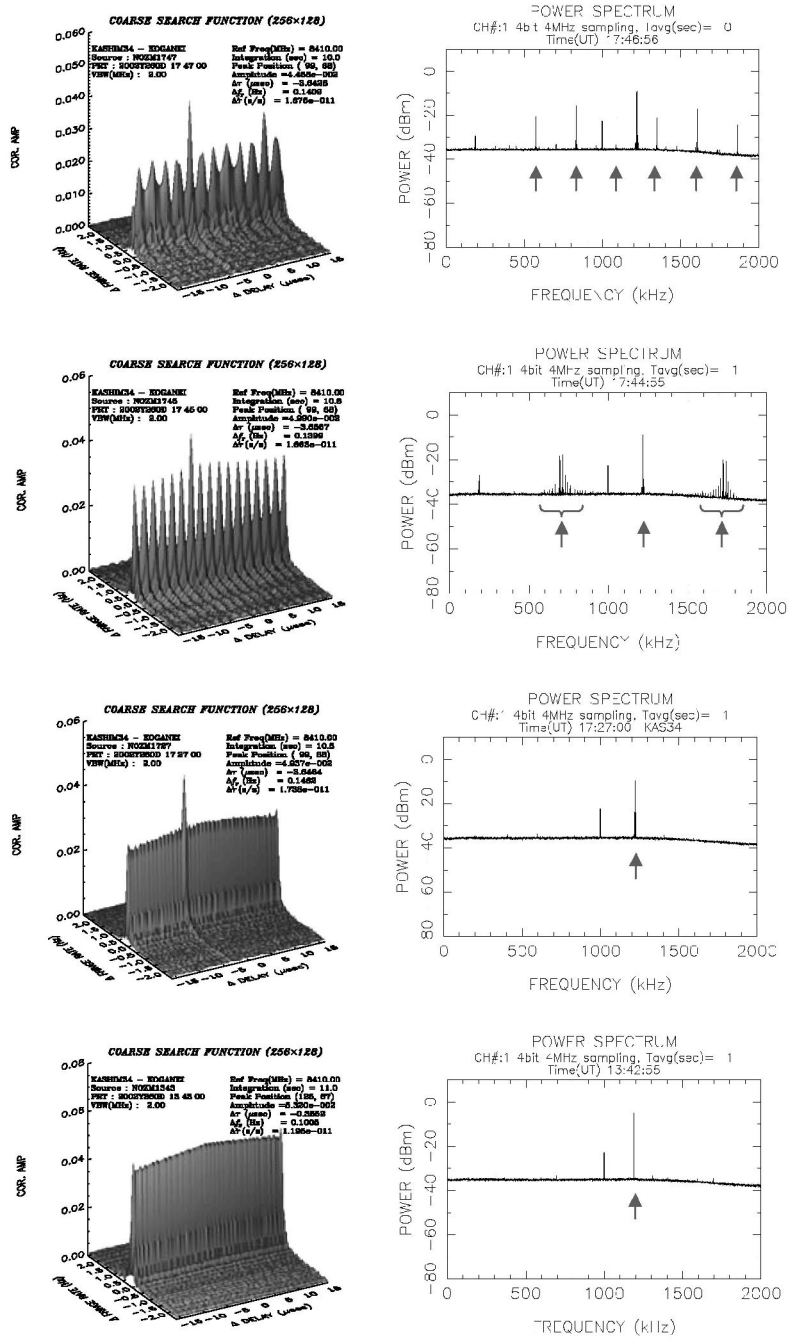


Fig.7. Fringes detected for NOZOMI. Group delays can be determined without ambiguity excepting the lowest panel case.

will remain. This is the idea of delta VLBI observation (e.g., Thornton and Border, 2003).

We have carried out a series of delta VLBI observations of GEOTAIL and NOZOMI using the K5-VSSP system since June, 2002 to learn characteristics of satellite downlink signals, to investigate the feasibility of delay measurements, and to utilize them for an actual orbit determination.

Two input-channels of a K5-VSSP are assigned for S and X band downlink signals. Observations are made with a sampling frequency of 4 MHz and an A/D resolution of either 4 bits or 2 bits. Observed data are FTPed to Kashima and are cross-correlated by PC. At present it is possible to obtain the delay time that is a VLBI observable to use for satellite positioning within a few days after observations. Fig. 5 shows the location of stations participated in the observations. Automatic observation control software

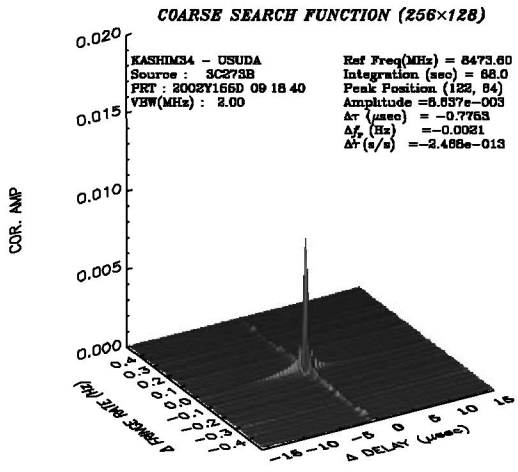


Fig.8. Fringe for QSO (3C273B).

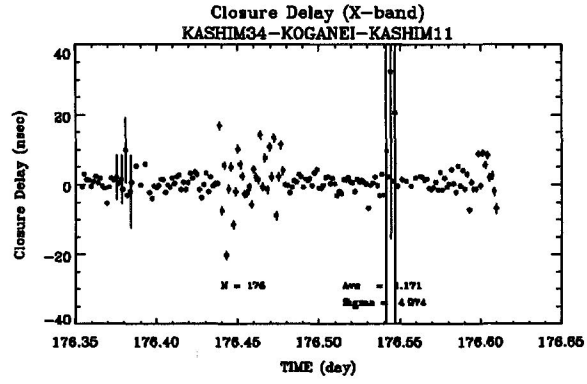


Fig.9 An example of closure delay test for GEOTAIL observations. Observations were carried out on June 25, 2002 at three stations, Kashima34, Kashima11, and Koganei11. Data with large error bars represent quasar observations. Although data are largely scattered around 176.45 days, standard deviation of total data is well less than 10 nsec.

was improved through a series of actual observations. Data reduction software was also developed and has been improved.

3.1 Fringes for different telemetry modes

In the differential one-way range (DOR) technique developed by JPL (Thornton and Border, 2003), telemetry signal is correlated with a computer generated model tone-signal to obtain the phase of each tone signal at each station. Then phases are differentiated between stations to get the cross correlation phase analogous to a quasar measurement. Delay is obtained as the slope of the phase versus frequency line. Unlike the DOR technique, we adopt the correlation processing as same as for a quasar processing, i.e., two station signals are directly correlated each other to obtain time delay. Thus the ordinary telemetry signal modulated by a complicated way is applicable for our delta VLBI observation. This increases the flexibility in a VLBI observation schedule because any special arrangement for telemetry mode is not necessary with a satellite control organization.

First we examined the possibility of group delay measurements by receiving telemetry signals. Then we investigated how the telemetry mode affects the delay measurements. Observed fringes (coarse delay search function) are summarized in Fig.6 for GEO-TAIL and Fig.7 for NOZOMI with the spectrum of telemetry signals.

As shown in the figures, group delay can be obtained without ambiguities when telemetry signal is modulated with any data. However ambiguities appear when the telemetry mode is in the range measurement mode or non-modulated mode (lowest panels of Figs 6 and 7), because insufficient number of tones (or lack of continuum component) on the frequency domain results in multiple peak structures on the time domain. For comparison sake, an example of fringe for a quasar (3C273B) is shown in Fig 8. Narrower peak in delay domain compared with the case of telemetry signals reflects the difference in the spectrum broadness, i.e., quasar has wider spectrum than telemetry's. We made a closure delay test to evaluate the accuracy of observed group delays. In the case shown in Fig.9 a root mean square error of a few nanoseconds was attained with coarse delay measurements. Nanosecond accuracy is obtained for other cases and it seems to be a typical accuracy for our group delay measurements. Since this accuracy is ten times or more badly than the accuracy for quasar delay measurements (about 0.1nsec or better is easily achieved), the total positioning accuracy is strongly influenced by the error of spacecraft-delay measurements. An error of 1 nsec in spacecraft delay measurements on a 100 km baseline corresponds to an angle resolution of 3 μrad.

3.2 Fringe detection for extremely weak telemetry signals

The NOZOMI is the first Japanese Mars orbiter, and was launched on July 4, 1998. It was planned to arrive Mars in 1999 by using two-times lunar swing-bys and a powered earth swing-by. However due to malfunction of a thruster valve during the powered earth swing-by and maneuvers to recover the right trajectory to Mars, enough fuel is not left to inject NOZOMI into a Mars orbit. The NOZOMI team at ISAS

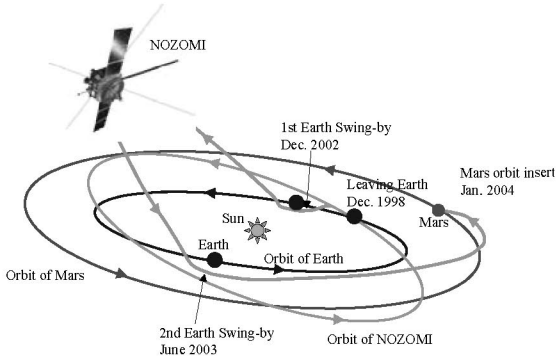


Fig.10. Trajectory of NOZOMI (after http://www.isas.ac.jp/e/enterp/missions/nozomi/traject.html)

found a new trajectory to Mars available to inject NOZOMI into a Mars orbit. Hence the orbit insertion scheduled in 1999 was abandoned, and it is now scheduled early in 2004 after two more earth swing-bys (Fig.10). During the two earth swing-bys telemetry signal intensity was supposed to become very weak due to the bad geometrical relation between spacecraft attitude and earth position, i.e., the high-gain antenna of NOZOMI does not point the Earth during this period. We carried out VLBI observations during this period to assess the sensitivity of our system and demonstrate the measurement accuracy if possible.

The NOZOMI moved to the high declination area after the first earth swing-by in December, 2002. Signal intensity became so weak from January to April, 2003 that integration time of 30 minutes or more was required to secure sufficient SNR. New fringe search software was hence developed to integrate the data over 30 minutes or more. An idea searching the acceleration of fringe rate was introduced into the new software to make possible to get fringes under insufficient a priori value. This new fringe search worked well as demonstrated in Fig 11.

4 Conclusions

We have carried out a series of delta-VLBI observations of GEOTAIL and NOZOMI using IP-VLBI system since June, 2002 to learn characteristics of satellite downlink signals and to investigate the feasibility of delay measurements. Results show that most of the cases we can determine group delays for telemetry signals with an accuracy of a few nanoseconds without any ambiguities. This accuracy will give an angle resolution of a few μrad on domestic baselines in Japan (about 100km or so), but it is insufficient resolution compared with a requirement of 0.5 μrad for NOZOMI navigation. Phase delay measurements under development will increase the measurement

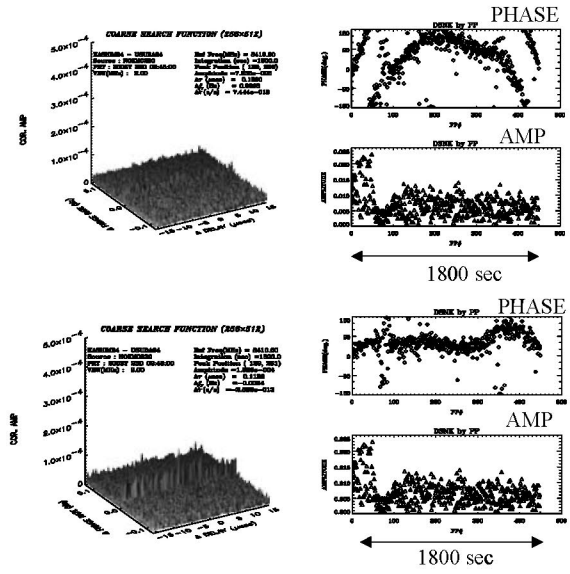


Fig.11. A normal fringe search for 30 minute scan data (upper two panels) and an improved fringe search (lower two panels).

accuracy by an order or more, but ambiguities must be solved for the orbital determination. We intend to develop orbital determination software that solves the ambiguity and the position simultaneously.

Acknowledgments. GEOTAIL and NOZOMI VLBI observations were carried out in collaboration with the ISAS. We would like to express our appreciation to M. Yoshikawa and his colleagues at ISAS for supplying the orbital information of satellites necessary for VLBI observations and for the arrangement of Usuda 64-m antenna for VLBI observations.

References

Kondo T, Koyama Y, Nakajima J, Sekido M, Ichikawa R, Kawai E, Okubo H, Osaki H, Kimura M, Ichikawa Y, and GALAXY team (2002) Real-time Gigabit VLBI System and Internet VLBI System, IVS 2002 General Meeting Proceedings, pp142-146

Koyama Y, Kurihara N, Kondo T, Sekido M, Takahashi Y, Kiuchi H, and Heki K (1998) Automated geodetic Very Long Baseline Interferometry observation and data analysis system, Earth, Planets and Space, 50, pp 709-722

Nishida A (1994) The GEOTAIL mission, Geopys. Res. Lett., 21, pp 2871-2873

Thornton C L and Border J S (2003) Radiometric tracking techniques for deep-space navigation, Deep-space communications and navigation siries Monograph 1, John Wiley & Sons Inc

Yamamoto T and Tsuruda K (1998) The PLANET-B mission, Earth Planets Space, 50, pp 175-18

Rapid Turnaround EOP Measurements by VLBI over the Internet

Yasuhiro Koyama, Tetsuro Kondo, Hiro Osaki
Kashima Space Research Center, Communications Research Laboratory,
893-1 Hirai, Kashima, Ibaraki 314-8501, Japan

Alan R. Whitney and Kevin A. Dudevoir
Haystack Observatory, Massachusetts Institute of Technology,
Route 40, Westford, MA 01886-1299, USA

Abstract. At present, promptness of the data processing of the international Very Long Baseline Interferometry observations is limited by the time required to ship observation tapes or disks from radio telescope sites to a correlation processing site. It can take more than two weeks to transport all the data tapes to the correlation processing site in typical observation sessions for Earth Orientation Parameter measurements. The e-VLBI technique uses high speed communication networks to transport observation data electrically and it has a possibility to dramatically improve the promptness of the data processing and hence the accuracy of the near real-time prediction of the Earth Orientation Parameters. To realize the e-VLBI observations by using a shared network environment, PC based data acquisition systems and data processing systems have been developed. By using the newly developed system, international e-VLBI sessions were performed twice and a rapid estimation of UT1-UTC in less than one day after the observations was demonstrated by the second session performed on June 27, 2003.

Keywords. VLBI, EOP, UT1, e-VLBI

1 Introduction

Currently, minimum delay time to obtain results from global geodetic Very Long Baseline Interferometry (VLBI) observations is about two weeks for typical one day experiments with several observing stations. This delay becomes shorter for so called intensive sessions where only one baseline is used and the duration is only about 1.5 hours, but it is still about one week or more. The purpose of the intensive sessions is to estimate the value of UT1-UTC by geodetic VLBI observations, and the sessions are performed four or five times every week by using either the KokeePark-Wettzell baseline or the Tsukuba-Wettzell baseline. Since only the geodetic VLBI technique can precisely measure UT1-UTC values independently, the results of the intensive sessions are essential information to generate predicted values of UT1-UTC (McCarthy, 1993). Since UT1-UTC randomly varies reflecting the dynamic phenomenon on the surface of the Earth and the interior of the Earth, the accuracy of the predicted UT1-UTC values deteriorates with time. Therefore, it is expected that the accuracy of the predicted UT1-UTC can be improved by shortening the delay time from the observations to the data processing of geodetic VLBI. The report of the IVS Working Group 2 for Product Specification and Observing Program (Schuh, et al., 2002) pointed out it is very important to shorten this processing time delay to improve the accuracy of the Earth Orientation Parameters (EOP) including UT1-UTC. The report also set the goal of the IVS observing program to make the time delay from observing to product less than one day by the year 2005.

To achieve this goal, it is necessary to transfer observation data electrically over the high speed global communication network from the observing sites to the correlator site. This form of VLBI is called e-VLBI to distinguish it from the conventional VLBI which uses magnetic tapes to record observed data. The first operational e-VLBI system was realized in the Key Stone Project of the Communications Research Laboratory (CRL) by using dedicated high speed research networks. In the system, four VLBI observing sites were connected with 2.4 Gbps Asynchronous Transfer Mode (ATM) network lines and the one day geodetic VLBI sessions were performed every two days. Although the baselines were relatively short to

estimate EOP values precisely, rapid turnaround EOP measurements less than one hour were technically demonstrated with the system (Koyama, et al., 1999).

After the routine observations of the Key Stone Project were terminated in 2001, new system developments were initiated to realize e-VLBI over the shared high speed networks. In this paper, newly developed observation and data analysis systems for e-VLBI will be introduced and the results of test observations by using the systems will be reported.

Table 1. Specifications of the IP-VLBI (VSSP) board.

Reference Signals	10MHz (+10dBm) and 1PPS
# of Input Ch.	1 or 4
A/D bits	1, 2, 4, or 8
Sampling Freq.	40kHz, 100kHz, 200kHz, 500kHz, 1MHz, 2MHz, 4MHz, 8MHz, or 16MHz
Bus Interface	PCI
OS	FreeBSD, LINUX, or Windows2000

2 System Developments

IVS Technology Development Centers including Haystack Observatory and CRL have been concentrating their efforts to realize e-VLBI in the global geodetic VLBI observations. At CRL, the K5 VLBI system has been developed based on the UNIX PC systems while the Mark-5 VLBI system has been developed at Haystack Observatory (Whitney, 2003). The K5 VLBI system (Fig. 1), also called the Versatile Scientific Sampling Processor (VSSP), is designed to perform real-time VLBI observations and correlation processing using Internet Protocol over commonly used shared network lines. It consists of four UNIX PC systems. Each UNIX PC system has one IP-VLBI data sampling board (Fig. 2), or also called as a VSSP board, on its PCI interface bus. Table 1 lists the specifications of the board. The board can sample 4 channels of base-band signals at various sampling rates ranging from 40kHz to 16MHz. The timing of the sampling is controlled by the provided 10MHz and 1PPS reference signals so that precise timing information can be reproduced from the sampled data. Quantization bits can be set from 1, 2, 4, and 8. Because the board has these many sampling modes, it has many possibilities to be used not only for VLBI observations but also for various other scientific researches which require precise timing information in the data.

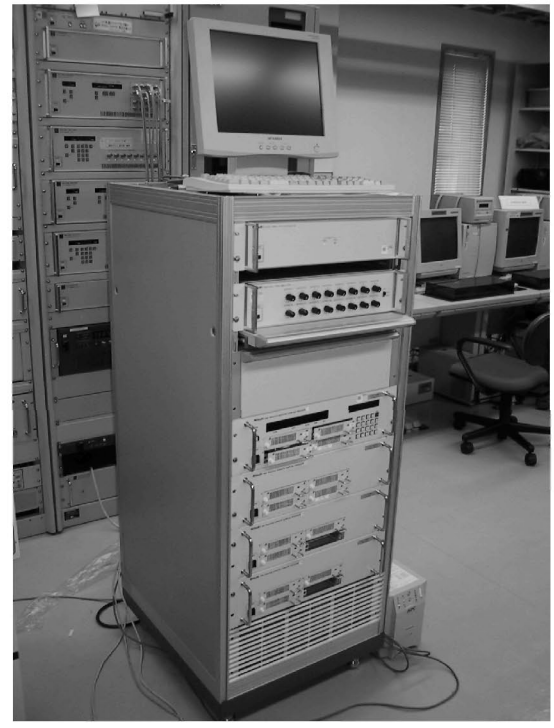


Fig. 1 A picture of the prototype K5 VLBI system (VSSP).

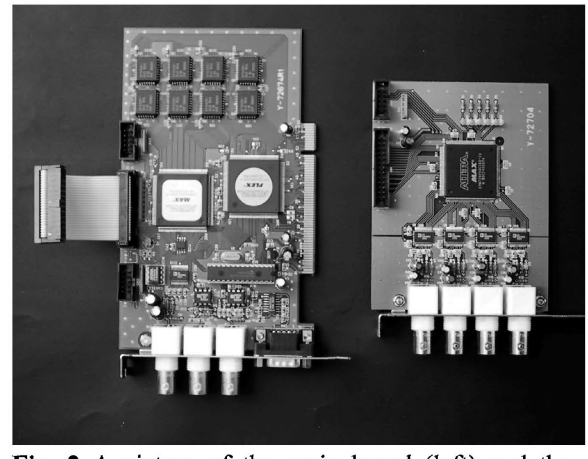


Fig. 2 A picture of the main board (left) and the auxiliary board (right) of the IP-VLBI (VSSP) board. Two boards are connected by the cable attached to the main board in the picture.

Device driver software of the board has been developed on LINUX, FreeBSD, and Windows2000 operating systems, and FreeBSD is used in the

prototype K5 data acquisition terminals. Two prototype K5 data acquisition systems have been configured. Four PC systems are mounted in the lower part of the 19-inch standard rack. A signal distributor unit for 1-PPS and 10 MHz signals and a 16-channel base-band signal variable amplifier unit are mounted in the upper part of the rack. The monitor and the keyboard on the top of the rack are connected to the four PC systems by using a four-way switch. Each PC system is equipped with four removable hard disk drives with data capacity of 120 GBytes each. The sampled data can be transferred to the network by using TCP/IP protocol or can be recorded to internal hard disks as ordinary data files. The maximum recording speed is currently restricted by the speed of the CPU and the speed of the PCI internal bus. Currently, the total recording speed of 512 Mbps has been achieved. It can be expected to record data up to 1024 Mbps by using a faster PCI bus and a faster CPU in the near future. To process the data sampled with the K5 data acquisition system, a software correlation processing program is also under development on FreeBSD PC systems. The correlation processing program receives data from K5 data acquisition systems over the network using TCP/IP protocol and then calculates cross correlation functions in real-time. It can also read data files on internal hard disks. These capabilities allow us to transfer observed data in real-time if the connecting network is fast enough, or in near real-time if data buffering is required. Since easily re-writable software programs and general PC systems are used, the processing capacity and the function of the correlator can be easily expanded and upgraded.

3 Network

Fig. 3 and Fig. 4 show the schematic route of the high speed network used in the e-VLBI test experiments. Two VLBI stations, one is 34-m and the other is 11-m, at Kashima and 11-m VLBI station at Koganei are connected to the Galaxy network under the collaboration between CRL and NTT laboratories. The maximum data rate of the Galaxy network is 2.4 Gbps and it supports ATM. At Kashima, an IP interface unit is connected to the ATM switch and it provides IP connection at the maximum data rate of 622 Mbps. The Galaxy network is then connected to the Super-SINET network at the National Institute for Informatics at Hitotsubashi. In the United States, the 18-m Westford VLBI station is connected to the Abilene network operated by the Internet2 consortium. The speed of the backbone of the Abilene network is now at 10 Gbps. The Abilene network and the Super-SINET network are connected at New York with a trans-Pacific link. The connection between the two networks is currently at 622 Mbps. There is also another route which uses GEMnet between Galaxy and the Abilene network. There is a plan to upgrade the speed of the GEMnet route to 155 Mbps in the near future, but the current speed is limited to 20 Mbps because of a bottleneck between Musashino and the Tokyo access point. During the test session on March 25, the route of Super-SINET was used.

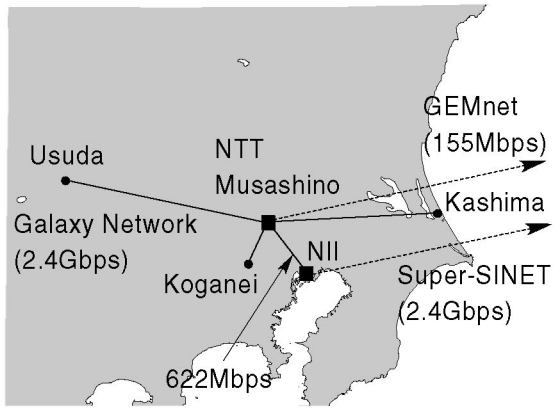


Fig. 3 Configuration of the high speed network in Japan.

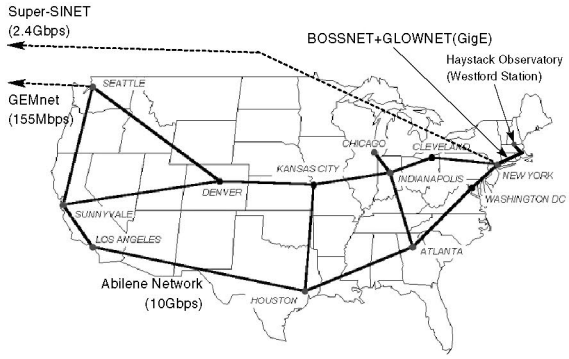


Fig. 4 Configuration of the high speed network in the United States.

4 Experiments

4.1 Kashima-Koganei Experiment

After completing two prototype K5 VLBI data acquisition terminals, a 24-hour geodetic e-VLBI

experiment was performed using two 11-m antennas at Kashima and Koganei from January 31 to February 1, 2003. Eight channels were assigned to both X-band and S-band and the total data rate was 64 Mbps. Observed data were stored in the internal hard disks as the data files. The data files were read by the software correlation program and the cross correlation processing was performed after all the observations finished. Then the bandwidth synthesis processing was performed and the obtained data were analyzed by CALC and SOLVE software developed by Goddard Space Flight Center of the National Aeronautics and Space Administration. During the observations, tape-based K4 data acquisition systems were used at both sites in parallel to compare the results. The data obtained with the K4 systems were processed with the K4 correlator at Kashima and analyzed similarly with the data obtained with the K5 systems. The results are compared in Table 2.

Table 2. Comparison of baseline lengths estimated from the data obtained with K4 and K5 systems.

	Baseline Length (mm)	RMS Residual Delay (fsec/sec)	Rate (psec)
K4	109099657.0 ± 6.7	76	136
K5	109099641.2 ± 3.2	33	92

From the comparison, the estimated baseline lengths are consistent with each other within twice the estimated uncertainties. In addition, the comparison of the RMS residuals of delay and delay rates suggests the performance of the K5 systems is better than the K4 systems. Part of the reason for the improvement is that the phase calculation of the phase calibration signals by the software correlation processing uses precise formulas whereas the K4 hardware correlator uses a three level approximation for sine and cosine functions for faster processing and to make the design of the hardware correlator very simple.

4.2 Kashima-Westford Experiments

A test e-VLBI session was performed for two hours from 16:00 UT on March 25, 2003 with the Kashima-Westford baseline. The 34-m antenna VLBI station at Kashima and the 18-m antenna station at Westford were used for the observations. This was the fourth test in the series of e-VLBI test observations. During the previous tests, successful detections of the fringes from the e-VLBI observations were demonstrated and the software developments have been continued with the data sets obtained.

At the Kashima station, the K5 VLBI system was used to record the observed data. At the Westford station, the Mark5 VLBI system developed by Haystack Observatory was used to record the observed data. The observed data were recorded to internal hard disks at each site and transferred to Kashima and Haystack Observatory after the observations by using FTP. For this purpose, a program was executed to extract the data from the Mark5 and to generate data files. At Haystack Observatory, K5 data files were converted and recorded to the Mark5 system. After the conversion, data recorded at Kashima and at Westford were processed for cross correlation processing using the Mark4 correlator at Haystack Observatory. At Kashima, the Mark5 data files were converted to K5 file format and then data recorded at Kashima and at Westford were processed by using the software correlator on the K5 VLBI system. After the cross correlation processing, bandwidth synthesis processing was done and two database files (one for X-band and another for S-band) were generated for data analysis. The sampling rate was hence 56Mbps. The observations were performed with 14 channels (8 for X-band and 6 for S-band) and 2MHz for each channel. The file size of the Mark5 system was checked from the files extracted from the Mark5 disks. Since there are 16 input channels both in K5 and Mark5 systems, two channels were used as redundant channels.

Table 3. Estimated value of UT1-UTC from the Kashima-Westford e-VLBI session and reported values from IERS Bulletin B 183, May 2, 2003.

	Epoch (UT)	UT1-UTC (μsec)
e-VLBI	20:00 on Mar. 25	-338727.0±23.9
IERS	00:00 on Mar. 25	-337951
IERS	00:00 on Mar. 26	-338610

After the data processing, CALC and SOLVE software were used to perform data analysis. In the estimation process, positions of both Kashima and Westford stations were fixed to the ITRF2000 values and the UT1-UTC was estimated along with

the clock offset, clock rate, and the atmospheric zenith delay. The estimated UT1-UTC value is shown in the Table 3 as well as the values published in the Bulletin B of the International Earth Rotation and Reference Systems Service (IERS). As shown in Table 3, UT1-UTC was estimated with uncertainty of 23.9 microseconds, which is comparable to the results from intensive sessions. The comparison of the estimated value with the results reported in IERS Bulletin B suggests there is a discrepancy between them. It is necessary to investigate the cause of the discrepancy. Especially, the consistency of the station coordinates and a priori information used in the data analysis has to be set carefully.

Table 4. Time sequence from the observations to the data analysis. Time is in Japanese Standard Time and start from 22:00 on June 27

Time	Event
22:00	Observations Start
00:00	Observations End
~04:20	File extraction and transmission From Kashima to Westford : 107Mbps (41.54GByte in 51m 35s) From Westford to Kashima : 44.6Mbps (41.54GByte in 2hr 04m 02s)
~08:10	File Conversion (Mark5 to K5)
~20:30	Software Correlation
~21:20	Bandwidth Synthesis Processing, Database Generation, Data Analysis

The next e-VLBI session was performed for two hours beginning at 13:00 UT on June 27, 2003. The observation mode, configuration of the observing systems, and the baseline were identical to the previous e-VLBI session performed in March, 2003. The purpose of the session was to demonstrate how fast the UT1-UTC can be actually estimated from the international e-VLBI session. Therefore, the file transfer servers as well as the network switches were carefully tuned to improve the file transfer speed over the network. Because the transmission delay over the network is quite large, about 200 msec, the default buffer size of the servers had to be increased from 64 kBytes to more than 4 Mbytes. As a result of this tuning, the file transfer speed was improved from about 2.2 Mbps to more than 100 Mbps. Table 4 shows the actual time sequence of the observations, file transfers, and data processing during the e-VLBI session. As shown in the table, the file transfer speed reached 107 Mbps in the direction from Kashima to Haystack Observatory. The opposite direction was not as fast, but the speed was about 45 Mbps. In total, the UT1-UTC was estimated within 21 hours and 20 minutes after the session finished. Thus the rapid estimation of the EOP in less than one day was successfully demonstrated by the international e-VLBI observations and data analysis.

5 Conclusions and Future Plan

Since file transfer speed became so fast, the longest time is required for data correlation processing. But since it is done by the software correlator, the time can be decreased easily by increasing the number of CPUs to be used in the correlation processing. File conversion is also a time-consuming processing, but this becomes unnecessary if K5 systems are used at both observing sites or if the file format between different systems were compatible. The discussions about the standard file format for e-VLBI have just begun, and it is expected that the file conversion processing will become unnecessary in the near future. By combining these efforts, it will become feasible to estimate UT1-UTC within a few hours after the observations without major difficulties.

The successes of the series of e-VLBI experiments have many important meanings. Two different recording systems have been developed by two independent teams and these two systems were used but the compatibilities between the two different systems were easily achieved by preparing file conversion software programs. The establishment of the file format standard for e-VLBI will further improve the compatibilities among different systems. It was also demonstrated that the international network connection has been drastically improved recently and the e-VLBI observations with intercontinental baselines are becoming realistic. Finally, the success of the bandwidth synthesis and the analysis of the from the data obtained with the K5 systems showed the improved performance of the prototype K5 VLBI data acquisition terminal compared with the conventional systems.

In the future, we are planning to continue similar e-VLBI sessions until the e-VLBI observations become reliable and robust. Since the trans-Pacific network has more capacity than we have used, it is expected that we can still increase

the file transfer speed by investigating the bottleneck of the network route. In parallel, we are planning to develop a CPU array system for distributed fast and effective correlation processing. The real-time data transfer and processing software programs will also be developed to improve further the promptness of obtaining results from geodetic VLBI observations. After these system developments, it will become possible to perform routine e-VLBI sessions with international baselines for rapid turnaround EOP measurements. To realize e-VLBI observations with isolated radio telescope sites, satellite communications have great possibilities. If satellite communications are utilized in real-time VLBI applications, variable communication delay will become the most problematic factor. It is highly desired to pursuit the use of satellite communications in the e-VLBI experiments as soon as possible.

Acknowledgements

The authors would like to acknowledge many members of the Haystack Observatory for observations and data processing, Internet2 for the Abilene network, NTT corporations for GALAXY network, National Institute for Informatics for Super-SINET network, and two reviewers for careful review on this paper.

References

Koyama, Y.. R. Ichikawa, T. Otsubo, J. Amagai, K. Sebata, T. Kondo, and N. Kurihara (1999), Recent Achievements in Very Long Baseline Interferometry, J. Comm. Res. Lab., No. 2, Vol. 46, pp. 253-258.

McCarthy, D. D. (1993), The Use of Crustal Dynamics Project Data to Predict the Orientation of the Earth, in Contributions of Space Geodesy to Geodynamics: Earth Dynamics, by D. E. Smith and D. L. Turcotte (eds.), Geodynamics Series Vol. 24, American Geophysical Union, pp. 71-76.

Schuh, H., C. Patrick, H. Hase, E. Himwich, K. Kingham, C. Klatt, C. Ma, Z. Malkin, A. Niell, A. Nothnagel, W. Schlüter, K. Takashima, and N. Vandenberg (2002), Final Report, IVS Working Group 2 for Product Specification and Observing Program, in 2001 Annual Report, International VLBI Service for Geodesy and Astrometry, by N. R. Vandenberg and K. D. Baver (eds.), NASA/TP-2002-00817-0, pp. 13-45.

Whitney, A. R. (2003), The Mark 5 VLBI Data System and e-VLBI Development, in 2002 Annual Report, International VLBI Service for Geodesy and Astrometry, by N. R. Vandenberg and K. D. Baver (eds.), NASA/TP-2003-211619, pp. 22-33.

Determination of tropospheric parameters within the new IVS Pilot Project

H. Schuh, J. Boehm
Institute of Geodesy and Geophysics,
University of Technology, Vienna, Gusshausstrasse 27-29, A-1040 Wien, Austria

G. Engelhardt
Federal Agency for Cartography and Geodesy, BKG, Leipzig, Germany

D. MacMillan
NASA Goddard Space Flight Center, GSFC, Greenbelt, U.S.A.

R. Lanotte
Centro di Geodesia Spaziale, CGS, Matera, Italy

P. Tomasi and M. Negusini
Istituto di Radioastronomia, CNR, Bologna, Italy

I. Vereshchagina and V. Gubanov
Institute of Applied Astronomy, IAA, St. Petersburg, Russia

R. Haas
Onsala Space Observatory, OSO, Onsala, Sweden

Abstract. In April 2002 the International VLBI Service for Geodesy and Astrometry (IVS) set up the Pilot Project - Tropospheric Parameters, and the Institute of Geodesy and Geophysics (IGG), Vienna, was put in charge of coordinating the project. Seven IVS Analysis Centers have joined the project and regularly submitted their estimates of tropospheric parameters (wet and total zenith delays, horizontal gradients) for all IVS-R1 and IVS-R4 sessions since January 1st, 2002. The individual submissions are combined by a two-step procedure to obtain stable, robust and highly accurate tropospheric parameter time series with one hour resolution. The zenith delays derived by Very Long Baseline Interferometry (VLBI) are compared with those provided by the International GPS Service (IGS). At sites with co-located VLBI and GPS antennas, biases are found between the Global Positioning System (GPS) and VLBI derived zenith delays, although short-term variabilities generally show a good agreement. Possible reasons for these biases are discussed. Additionally, the time series submitted by the individual Analysis Centers are compared to the combined IGS time series for the CONT02 campaign of 15 days of continuous observations in the second half of October 2002.

Keywords. VLBI, tropospheric path delays, co-located sites, CONT02.

1 Introduction

In the last few years, the collaboration between geodesy and meteorology/climatology has become more and more intensive. The Global Positioning System (GPS) has proved to be very important for meteorology, and because of the short delay between the GPS observations and the availability of tropospheric results, these can even be entered into weather forecast models. Tropospheric parameters determined by Very Long Baseline Interferometry (VLBI) are mainly useful for climatological studies, since at present the delay between a session and the availability of results is greater than one week. As there is a long history of consistently analyzed VLBI sessions since 1984, the wet zenith delays comprise accurate information about the long-term trends of precipitable water content above the VLBI sites. Furthermore, due to their high accuracy, the parameters derived by VLBI are of interest for the validation and calibration of parameters determined by GPS, water vapour radiometer (WVR) and other techniques.

At the 7th IVS Directing Board meeting in Tsukuba, Japan (Feb. 2002) it was decided that an IVS Pilot Project on Tropospheric Parameters should be set up and coordinated by the Institute of Geodesy and Geophysics, Vienna. This Pilot Project is a research and study project with a structure similar to the IVS Working Groups.

Table 1. IVS Analysis Centers taking part in the Pilot Project - Tropospheric Parameters.

	IVS Analysis Center
BKG	Fed. Ag. for Cartography and Geodesy, Germany
CGS	Centro di Geodesia Spaziale, Italy
CNR	Istituto di Radioastronomia, Italy
GSF	NASA Goddard Space Flight Center, U.S.A.
IAA	Institute of Applied Astronomy, Russia
IGG	Institute of Geodesy and Geophysics, Austria
OSO	Onsala Space Observatory, Sweden

After the call for participation by the IVS Analysis Coordinator in May 2002, seven IVS Analysis Centers agreed to take part in the PP. A Pilot Project Group was set up to coordinate all activities within the Pilot Project and to discuss all steps that should finally lead to operational products.

2 Submissions by the Analysis Centers and combination

Most of the Analysis Centers have provided their estimates of the tropospheric parameters beginning with January 2002. That allows the generation of a combined series since the start of the IVS-R1 and IVS-R4 sessions. These are 24 h experiments usually observed on Mondays and Thursdays, respectively. Total (hydrostatic plus wet) and wet zenith delays as well as gradient delay parameters are submitted by all Analysis Centers. GSF and IGG even apply a priori gradients calculated from numerical weather models. Most of the Analysis Centers use the CALC/SOLVE software package which is mainly maintained by NASA Goddard Space Flight Center, while IAA and IGG apply the QUASAR (Gubanov et al., 2002) and OCCAM (Titov et al., 2001) software, respectively. About half of the Analysis Centers fix station positions to the ITRF2000, and all use cutoff elevation angles at or below 5°. The Niell mapping functions (Niell, 1996) are used throughout - only IGG applies the isobaric mapping function for the hydrostatic part (Niell, 2001). The tropospheric parameters are computed for every full hour, i.e. in equidistant time intervals of 60 minutes, starting at the first integer hour of the session. If other time intervals are used for the computation (e.g., longer time intervals for the gradients), all parameters have to be referred to the same hourly instants. More details about the Pilot Project - Tropospheric Parameters, the Pilot Project Group and the submissions of the Analysis Centers are given by Schuh and Boehm (2003).

Table 2. Features of the submissions. Two Analysis Centers use a priori gradients; four Analysis Centers fix ITRF2000.

Analysis Center	a priori gradients	ITRF2000 fixed	software
BKG	no	yes	CALC/SOLVE
CGS	no	no	CALC/SOLVE
CNR	no	no	CALC/SOLVE
GSF	yes	no	CALC/SOLVE
IAA	no	yes	QUASAR
IGG	yes	yes	OCCAM
OSO	no	yes	CALC/SOLVE

The combination strategy for the total and wet zenith delays is a two-step procedure with a final data cleaning. The details can again be found in Schuh and Boehm (2003).

3 Internal accuracy and precision of the combined zenith delays

Only if the differences between the zenith delays by the different Analysis Centers were random, which is unlikely since they share data and in many cases software, would standard weighting schemes based on the rms-differences yield a reliable estimate of the accuracy of the combination. Therefore, the weekly biases between the Analysis Centers can only be a good criterion to evaluate the 'internal' (VLBI-) accuracy of the zenith delay combination. To evaluate the absolute accuracy of the zenith delay combination, comparisons with other techniques (e.g. radiosonde observations, GPS) have to be applied. Possible systematic errors due to the VLBI technique that might be inherent in the zenith delays submitted by all Analysis Centers might be:

- errors of the terrestrial reference frame (at least for those solutions where the ITRF2000 is fixed),
- errors of the mapping functions,
- unmodelled effects (atmospheric loading, antenna deformation, ..).

On the other hand, the precision (formal error) of the combination can be determined when the standard deviations at the hourly instants are evaluated (after removing the weekly biases between the time series).

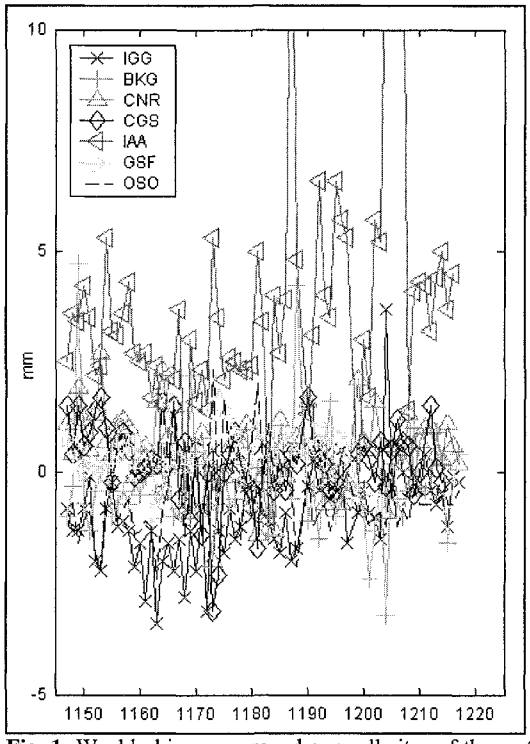


Fig. 1. Weekly biases averaged over all sites of the wet zenith delays since January 2002. The biases are within ±2 mm for most of the Analysis Centers. This figure is meant to provide a rough idea of the agreement between the time series. For figures of better quality, please refer to the webpage at http://www.hg.tuwien.ac.at/~ivstrop .

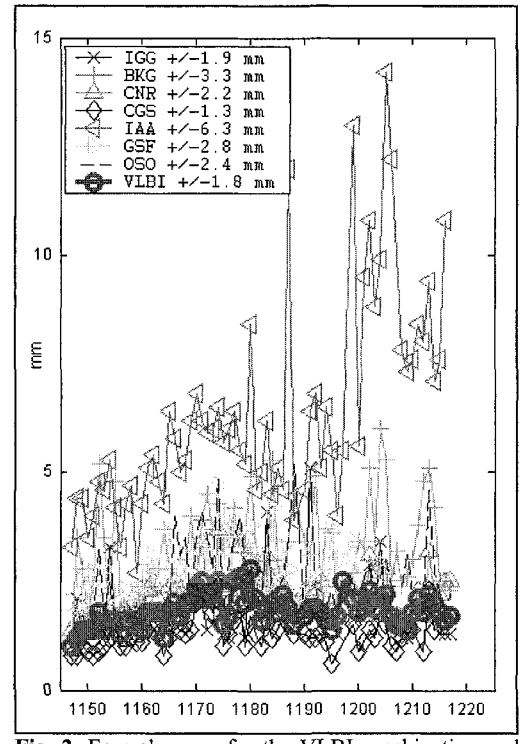


Fig. 2. Formal errors for the VLBI combination and standard deviations of the differences between the individual solutions and the combination for the VLBI Analysis Centers, respectively, since January 2002 (averaged over all sites). For figures of better quality, please refer to the webpage.

3.1 Internal accuracies

As can be seen in Figure 1, the weekly biases relative to the IVS combination of the wet (and total) zenith delays are within ±2 mm for most of the Analysis Centers. This indicates that - apart from systematic effects as described above - the accuracy of the zenith delays is at the 2 mm level, which is a mean value for all VLBI sites. Comparisons with other techniques, e.g. GPS, imply that the absolute accuracy is at the 5 mm level, which was found by Niell et al. (2001) who also concluded that the zenith delays derived by VLBI appear to be more accurate than estimates from GPS, WVR (water vapour radiometer) or radiosonde data.

3.2 Precision

The precision of the combination was calculated as the standard deviation at the hourly instants after removing the weekly biases (formal errors). Figure 2 shows the mean values (averaged per week and over all sites) of the formal errors of the combined VLBI solution (bold line) of the wet zenith delays. Additionally, the mean standard deviations of the differences of the individual time series relative to the combined VLBI solution are shown. Thus, the precision (formal error) of the combined VLBI wet (and total) zenith delays is ~ ±1.8 mm (see Figure 2). At the co-located GPS antennas and for identical epochs the formal errors for the total zenith delays as provided in the weekly files by IGS is ~ ±2.6 mm.

4 Comparison with tropospheric parameters determined by IGS

The IGS has produced tropospheric parameters for 150 IGS sites since 1997 (Gendt, 1996). This allows one to compare at co-located sites the total zenith delay combination derived by VLBI within the IVS-PP with those published by the IGS. The biases and standard deviations between the IGS and IVS time series are the parameters of interest. The height

differences between the VLBI and GPS stations are accounted for by using pressure data recorded at the VLBI stations and assuming an exponential decrease of pressure with height. As a rule of thumb, a height difference of 10 m corresponds to 3 mm differential hydrostatic delay. The differential wet delay is negligible, if not unpredictable.

4.1 IVS-R1 and IVS-R4 sessions

Table 3 shows the mean biases between the IGS and IVS time series and the standard deviations of the differences after removing these biases for all IVS-R1 and IVS-R4 sessions since 2002.0 .

Table 3. Co-located sites with VLBI and GPS antennas. The 4-letter IGS acronyms are given as well as the height differences (VLBI - GPS) between the antennas. The last two columns show the biases which remain after accounting for the height difference and standard deviations between the IVS and IGS time series for identical epochs.

site	IGS acronym	height diff. [m]	biases (IGS-IVS) [mm]	std.dev. ± [mm]
Algopark	algo	23.0	7.1	4.8
Fortleza	fort	3.3	13.5	9.6
Gilcreek	fair	13.1	4.2	3.7
Hartrao	hrao	2.3	5.2	8.1
Hobart26	hob2	24.9	3.2	7.4
Matera	mate	7.7	3.9	6.8
Medicina	medi	17.1	1.4	4.6
Nyales20	nyal	6.5	4.1	3.8
Seshan25	shao	8.2	1.5	6.0
Wettzell	wtzr	3.1	2.4	4.3
Onsala60	onsa	13.8	4.8	4.5
Westford	wes2	1.7	6.7	3.7
		mean:	4.7	5.8

Although the standard deviations between the IGS and VLBI time series are at the ±5 mm level or greater, it is noticeable that all mean values of the total zenith delays derived by GPS are larger than those derived by VLBI. The positive biases are between +1.4 mm (Medicina) and +13.5 mm (Fortaleza). This confirms results reported by Boehm et al. (2002). In addition to the systematic effects for VLBI described above, there are systematic effects for GPS observations as well:

- due to higher cutoff elevation angles applied in GPS processing (mostly larger than 10°) valuable information about the separation of zenith delays and station heights gets lost,
- multipath effects at low elevations,
- phase center variations of the antennas,
- errors of satellite ephemerides.

4.2 CONT02

All Analysis Centers also submitted their tropospheric estimates for CONT02. This was a special campaign with 15 consecutive 24 h experiments in the second half of October 2002 with the same eight stations taking part throughout. The goal of the CONT02 campaign was to acquire the best possible VLBI data over a two-week period to demonstrate the highest accuracy of which VLBI is capable. In this paper, CONT02 is used to confirm the conclusions from the IVS-R1 and IVS-R4 sessions and also to compare the time series from the individual Analysis Centers with the combined IGS time series. Figure 3 shows the zenith delay series for Gilcreek for the individual submissions, the IVS combination and the IGS solution. Table 4 shows that the biases between total delays derived from VLBI and GPS are at the +5 mm level, which is consistent with biases for the IVS-R1 and IVS-R4 sessions. This confirms the statement made in section 4.1 that the total delays derived from GPS are larger than those derived by VLBI even if the height difference is accounted for. Table 4 also shows the standard deviations between the total zenith delays from IGS and the individual solutions from the IVS Analysis Centers. It is obvious that for some stations there are problems with the submissions from IAA, which still remain to be clarified. Taking a look at the plots (see Figure 3 as an example) reveals that very often the IGS time series vary less than the VLBI series. The reason for this may be in part that the IGS combination series has a resolution of two hours whereas the IVS Analysis Centers are using time intervals of one hour or less. By that the VLBI results allow to detect short term variability of the tropospheric parameters. Although, for the comparison with IGS, only every second value was used, the larger variation is still inherent in the time series. Variability differences are also due to the differences in the constraints of the estimation models that were used.

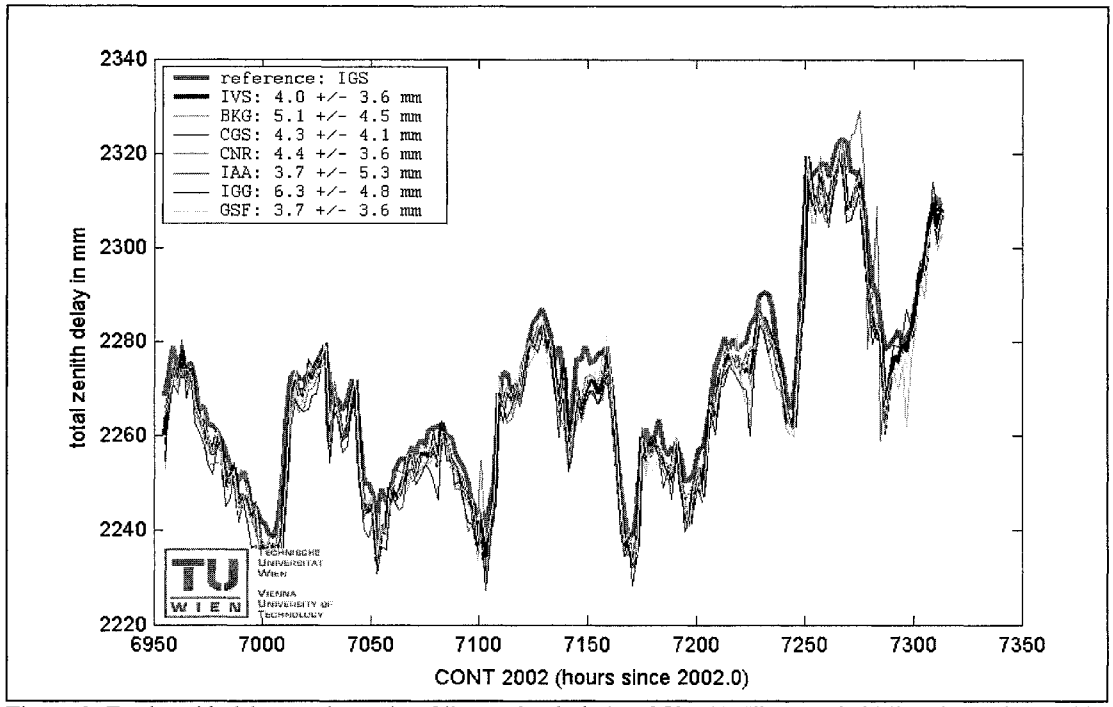


Figure 3. Total zenith delays at the station Gilmore Creek during CONT02. The grey bold line shows the combined GPS solution as published by the IGS The black bold line named IVS is the combined IVS time series. This figure is meant to provide a rough idea of the agreement between the time series. For figures of better quality, please refer to the webpage at http://www.hg.tuwien.ac.at/~ivstrop .

Table 4. Biases and standard deviations (after removing the biases) in mm between the combined IGS time series and the individual Analysis Centers of the IVS. The last two columns show the biases and standard deviation between the combined IGS and the combined IVS total zenith delays.

	BKG		CNR		CGS		IAA		IGG		GSF		OSO		IVS	
	bias	std	bias	std	bias	std	bias	std	bias	std	bias	std	bias	std	bias	std
Algopark	7,7	4,4	5,8	4,5	5,6	4,6	-22	45,0	7,0	4,8	4,1	3,8	7,5	5,2	6,0	4,0
Gilcreek	5,1	4,5	4,3	4,1	4,4	3,6	3,7	5,3	6,3	4,8	3,7	3,6	3,9	4,0	4,0	3,6
Hartrao	4,5	6,8	3,9	8,1	3,5	7,9	-21	50,9	3,4	6,8	4,4	5,3	3,3	7,5	3,8	6,2
Kokee	9,7	8,1	6,9	8,7	6,8	8,4	-11	69,4	11,4	7,8	9,5	6,3	7,9	8,3	8,8	7,1
Nyales20	4,3	5,4	1,3	4,0	1,9	3,8	-5,5	14,1	5,8	3,9	1,9	3,5	3,7	4,4	3,0	3,6
Wettzell	3,5	6,1	3,5	5,9	4,1	6,0	-2,4	6,7	4,6	5,8	3,2	4,6	1,5	5,8	3,7	5,2
Westford	9,2	4,9	7,0	5,0	7,0	4,8	4,8	5,5	9,3	4,4	6,8	4,4	8,4	5,1	7,9	4,3
Onsala	5,0	5,3	2,3	5,0	2,6	5,2	-0,7	12,7	6,9	4,3	4,1	4,6	3,8	5,0	4,2	4,4
mean:	**6,1**	**5,7**	**4,4**	**5,7**	**4,5**	**5,6**	**-6,7**	**26,2**	**6,8**	**5,3**	**4,7**	**4,5**	**5,0**	**5,7**	**5,2**	**4,8**

5. Results and conclusions

Based on rms-differences between the IVS and IGS combinations, the precision of tropospheric zenith delays from VLBI is at ~ ±3.5 mm. The Pilot Project has clearly shown that comparing and combining the results of several Analysis Centers which use different VLBI software or apply different analysis strategies allows

- feedback to be given to the individual Analysis Center in case of any problems,
- determination of stable, robust and highly accurate final IVS products with standard deviations that are usually significantly smaller than those of the individual submissions.

As this Pilot Project proved to be very successful, it was decided at the 9th IVS Directing Board meeting in Paris (April 2003) that the tropospheric parameters should become operational IVS products. IGG was asked to coordinate the project called IVS-TROP.

Zenith delays derived by VLBI can be compared to those derived by GPS. The agreement is generally rather good but the always positive biases between the GPS and VLBI time series at co-located sites need to be investigated in more detail.

The other field of application for zenith delays derived by VLBI is the contribution to climatological studies, at least when the time series cover a longer time interval. First results are published by Boehm et al. (2003), reporting that the trend in the wet zenith delays at Wettzell in the last two decades is ~ 0.83 mm/yr.

References

Boehm, J., H. Schuh and R. Weber (2002). Influence of tropospheric zenith delays obtained by GPS and VLBI on station heights. In: Proceedings of the IAG Symposium on Vertical Reference Systems, Cartagena, Colombia, Feb. 2001, ed. by H. Drewes, A. Dodson, L. Fortez, L. Sanchez, P. Sandoval, Springer Verlag Berlin-Heidelberg.

Boehm, J., H. Schuh, V. Tesmer and H. Schmitz-Hübsch (2003). Determination of tropospheric parameters by VLBI as a contribution to climatological studies, Oesterreichische Zeitschrift fuer Vermessung & Geoinformation (VGI), 1/2003, ed. by H. Schuh, F. Brunner, and H. Kahmen, 21 - 28.

Gendt, G. (1996). Comparisons of IGS tropospheric estimates. Proceedings IGS Analysis Center Workshop, 19-21 March 1996 Silver Spring, Maryland USA, ed. by R. E. Neilan, P. A. Van Scoy, J. F. Zumberge, 151-164.

Gubanov, V.S., I.A. Kozlova, I.F. Surkis and Yu.L. Rusinov (2002). Processing of VLBI observations: Software QUASAR, Parts 1-5. Comm. IAA RAS, 141-145 (in Russian).

International GPS Service (2003), http://igscb.jpl.nasa.gov/ .

International VLBI Service for Geodesy and Astrometry (2003), http://ivscc.gsfc.nasa.gov/ .

Niell, A.E. (1996). Global mapping functions for the atmosphere delay at radio wavelength, J. Geophys. Res., 101 (B2), 3227-3246.

Niell, A.E. (2001). Preliminary evaluation of atmospheric mapping functions based on numerical weather models, Phys. Chem. Earth, 26, 475-480.

Niell, A.E., A.J. Coster, F.S. Solheim, V.B. Mendes, P.C. Toor, R.B. Langley, C.A. Upham (2001). Comparison of Measurements of Atmospheric Wet Delay by Radiosonde, Water Vapor Radiometer, GPS, and VLBI, J. of Atm. and Ocean. Technology, 18, 830-850.

Schuh, H. and J. Boehm (2003). Status Report of the IVS Pilot Project - Tropospheric Parameters in International VLBI Service for Geodesy and Astrometry 2002 Annual Report, ed. by N. R. Vandenberg and K. D. Baver, NASA/TP-2003-211619, 13-21.

VLBIONOS - PROBING THE IONOSPHERE BY MEANS OF VERY LONG BASELINE INTERFEROMETRY

T. Hobiger, J. Boehm, H. Schuh
Institute of Geodesy and Geophysics,
University of Technology, Vienna, Gußhausstraße 27-29, A-1040 Wien, Austria

Abstract. In geodetic Very Long Baseline Interferometry (VLBI) the observations are performed at two distinct frequencies (2.3 and 8.4 GHz) in order to determine ionospheric delay corrections. This allows information to be obtained from the VLBI observables about the sum of electrons (total electron content - TEC) along the ray path through the ionosphere. Due to the fact that VLBI is a differential technique, only the differences in the behavior of the propagation media over the stations determine the values of the observed ionospheric delays. However, in a first simple approach, an instrumental delay offset per baseline shifts the TEC measurements by a constant value. This offset is independent of the azimuth and elevation of the observed radio source what allows separation of the ionospheric parameters for each station from the instrumental delay offsets per baseline in a least-squares adjustment. In first tests of this method Fourier coefficients up to the 4th order plus a constant value and a linear trend were estimated to represent the vertical TEC (VTEC). Slant TEC (STEC) values are converted into VTEC values by a mapping function. A disadvantage of this approach is the assumption that these values are assigned to the station coordinates but not to the geographical coordinates of the intersection point of the ray path and the infinitely thin ionospheric layer. The precision of the estimated values is about ±5 to ±7 TEC units (TECU). The results obtained from VLBI agree with a standard deviation of ±10 TECU with other techniques like GPS, and the differences rarely exceed 20 TECU. A second approach, developed at the TU Vienna, using piece-wise linear functions (VTM – Vienna TEC model) is also tested.

Keywords. VLBI, ionosphere, GPS, TEC

1 Introduction

The Earth's ionosphere is defined as that part of the upper atmosphere where free electrons occur in sufficient density to influence the propagation of electromagnetic radio frequency waves (Hargreaves, 1992). The ionization depends primarily on the Sun and its activity. Ionospheric structures vary strongly with time corresponding to the sunspot cycle and seasonal and diurnal cycles, and with geographical location (polar, auroral, mid-latitude and equatorial regions). Certain further ionospheric disturbances can be related to the Sun. The major part of the ionization is produced by solar X-ray and ultraviolet radiation and by corpuscular radiation from the Sun. The most noticeable effect is seen as the Earth rotates with respect to the Sun. Ionization increases in the sunlit hemisphere and decreases on the shadowed side. Although the Sun is the largest contributor to the ionization, cosmic rays are the source of a small contribution, too.

The ionosphere is a dynamic system depending on many parameters, including acoustic motions of the atmosphere, electromagnetic emissions, and variations of the geomagnetic field. Any atmospheric disturbance affects the distribution of the ionization, and, because of its extreme sensitivity to atmospheric changes, the ionosphere can be used as a sensor of atmospheric events. A description of the structure of the ionosphere, in terms of its vertical profile, its main geographic regions and its temporal variations is given by Hobiger et al. (2003).

2 Impact on space geodetic techniques

One of the problems for all space geodetic techniques operating with electromagnetic waves is the determination of the propagation velocity of the signals. If these waves propagated in vacuum, the traveled distance would be just the product of the propagation time between emitter and receiver and the speed of light in vacuum. When signals travel through the ionosphere, the interaction between the electromagnetic field and the free electrons influences both the speed and the propagation direction of the signal, an effect known as ionospheric refraction (Hartmann and Leitinger, 1984). The ratio between the propagation speed of a wave in vacuum c and the propagation speed in a given medium v_{ph} is known as the refractive index n

$$n = \frac{c}{v_{ph}} \tag{1}$$

The Appleton-Hartree theory (Hartmann and Leitinger, 1984) allows the calculation of the refractive index for a single wave which propagates through a plasma (= ionized medium). By neglecting higher order terms we obtain an approximation for n (see Hobiger et al., 2003).

$$n \approx 1 - \frac{40.28\, N_e}{f^2} \tag{2}$$

where f is the carrier frequency expressed in Hz and N_e is the free electron density in the medium. Using this result the group refractive index n_g can be determined by

$$n_g \approx 1 + \frac{40.28\, N_e}{f^2} \tag{3}$$

Now we can calculate the carrier phase delay, d_{ph}, to obtain the influence of the ionized medium on the propagation delay, expressed in SI units (meters)

$$d_{ph} = \int_S (n_{ph} - 1)\,dS = -\frac{40.28}{f^2}\int_S N_e\,dS = \\ = -\frac{40.28 \cdot 10^{16}}{f^2}\,STEC \tag{5}$$

The integral of the electron density along the signal path S is usually called $STEC$ (Slant Total Electron Content). This quantity can be interpreted as the total amount of free electrons in a cylinder with a cross section of 1 m^2 of which the axis is the slant signal path. $STEC$ is measured in Total Electron Content Units (TECU), which is equivalent to $10^{16}\,electrons \cdot m^{-2}$. The effect of the ionized medium on group propagation can be expressed by

$$d_g = \int_S (n_g - 1)\,dS = \frac{40.28}{f^2}\int_S N_e\,dS = \\ = \frac{40.28 \cdot 10^{16}}{f^2}\,STEC \tag{6}$$

The last two expressions show how the electron content in the ionosphere can influence measurements ranging from outer space to Earth-based stations. If the behavior of the ionosphere is known, these effects can be computed and can be used to correct measurements on radio frequencies. If ionospheric parameters are not available, observing at two different radio frequencies allows the elimination of the first order ionospheric influences.

3 A new field of research - ionospheric investigations by geodetic VLBI

The project 'VLBIonos', supported by the Austrian Science Fund (FWF), started on March 1st, 2003, and aims at the investigation of the ionosphere using geodetic VLBI. VLBI observations are performed at two different frequencies (2.3 and 8.4 GHz, S- and X-band) in order to determine the ionospheric delay. As shown by Kondo (1991) this information can be used to model the ionosphere above each station. It has to be mentioned that only the differences in the ionospheric delay between the two stations are measured. Nevertheless it will be possible to determine absolute values at each station without any external information or constraints, if we neglect horizontal variations in the ionosphere above the sites. Constant instrumental offsets at each station bias these measurements, which leads to

$$\tau_{model}(t) = \tau_{ion,1}(t) - \tau_{ion,2}(t) + \tau_{offset,1} - \tau_{offset,2} \tag{7}$$

The ionospheric delay $\tau_{ion}(t)$ at X-band (f_x) over station i can be modeled as

$$\tau_{ion,i}(t) = \frac{1.34 \cdot 10^{-7}}{f_x^2} \cdot S(E_i) \cdot VTEC_i(t) \qquad (8)$$

where

$$S(E_i) = \frac{1}{\cos\left\{\arcsin\left[\frac{R \cos E_i}{R+h}\right]\right\}} \qquad (9)$$

and $VTEC_i(t)$ represents the vertical TEC value at the intersection point of the ray path with the infinitesimally thin ionospheric layer assumed to be at the height h. The radius of the (spherical) Earth is abbreviated with R and E_i stands for the elevation angle at station i. Under the assumption that horizontal gradients in the ionosphere can be neglected within a range of about 300 kilometers, $VTEC_i(t)$ can be assigned to the geographical coordinates of the VLBI station. In our first investigation we concentrate on modeling the behavior of the ionosphere over the stations by two different approaches.

- Model proposed by Kondo (1991)

In this model $VTEC_{Kondo}(t)$ for station i and time t is calculated as suggested by Kondo (1991) using a constant offset, sine and cosine functions, and a daily rate

$$VTEC_{Kondo,i}(t) = \\ = a_{i0} + \sum_{k=1}^{4}\left[a_{ik}\cos\left(\frac{kt\pi}{12}\right) + b_{ik}\sin\left(\frac{kt\pi}{12}\right)\right] + c_i t \qquad (10)$$

- Vienna TEC Model (VTM)

In this model $VTEC_{VIENNA}(t)$ is calculated as a piece-wise linear function

$$VTEC_{VIENNA,i}(t) = offset_i + rate_{i1}(t_1 - t_0) + \\ + rate_{i2}(t_2 - t_1) + \ldots + rate_{in}(t_n - t) \qquad (11)$$
$$\text{and } t \le t_n$$

After calculating the partial derivatives

$$\left(\frac{\partial VTEC_{Kondo,i}}{\partial a_{ij}}, \frac{\partial VTEC_{Kondo,i}}{\partial b_{ij}}, \frac{\partial VTEC_{Kondo,i}}{\partial c_i}\right. \text{ or}$$
$$\left.\frac{\partial VTEC_{VIENNA,i}}{\partial offset_i}, \frac{\partial VTEC_{VIENNA,i}}{\partial rate_{ij}}\right) \text{ a least-squares}$$

adjustment allows separation of the VTEC values from the constant instrumental offset $\tau_{offset,i}$. This is possible because observations are performed at different elevation angles which prevents the design matrix from becoming singular. The VTM approach also includes constraints on the VTEC rate of about ±30 TECU per hour for the following two reasons: to get physically reasonable values and to get a non-singular design matrix even if there are gaps in the data.

4 Results

4.1 Ionospheric values and maps derived by GPS

Data relevant to this work stored in IONospheric Exchange format (IONEX) (see Schaer et al., 1998) on the IGS web server are provided by the following analysis centers:

- Center for Orbit Determination in Europe (CODE), University of Berne, Switzerland,
- Geodetic Survey Division of Natural Resources Canada – formerly Energy, Mines and Resources Canada (EMR), Ontario, Canada,
- European Space Operations Centre (ESOC) at the European Space Agency (ESA), Darmstadt, Germany,
- Jet Propulsion Laboratory (JPL), Pasadena, U.S.A.,
- Group of Astronomy and Geomatics, Universidad Politecnica de Catalunya (gAGE/UPC), Barcelona, Spain.

So far (summer 2003), a combined series of these GPS analysis centers does not exist. Additionally, regional ionospheric models were derived in this study (i.e. at the Institute of Geodesy and Geophysics (IGG) of TU Vienna) from GPS data using BERNESE software. For this purpose data from five to ten uniformly distributed stations around the 'center station' were taken to calculate a model represented by spherical harmonics. If the VLBI station is close to the center of the area covered by the GPS sites a better agreement between the two techniques can be expected than with IONEX because no interpolation of gridded data is necessary. In section 4.3 these regional GPS solutions will be referenced as GPS-IGG/BERNESE.

4.2 VTEC values derived from VLBI NEOS-A sessions

Figure 1 shows VTEC values determined by VLBI from the NEOS-A session on January 16th, 2001 over the station Fortaleza (thick lines) compared to VTEC numbers derived by GPS (thin lines). Fortaleza is located close to the equator and shows high maxima and a strong variability of the VTEC numbers during 24 hours. The results of both VLBI approaches (Kondo approach and VTM approach) agree quite well between each other and also with the solutions of the different IGS analysis centers. Sometimes the differences between the individual GPS solutions are even bigger than the differences from VLBI results what might depend on the various models chosen and the constraints set in the GPS analyses. In figure 2 the amplitudes

$$A_{ik} = \sqrt{a_{ik}^2 + b_{ik}^2} \tag{12}$$

of the harmonic terms are plotted i.e., for periods of 24h, 12h, 8h, and 6h, for all stations involved in this experiment according to the Kondo approach. The error bars show the precision of each amplitude A_{ik} and the x-axis corresponds to the geographical latitudes of the VLBI stations. It is obvious that stations which are closer to the equator show larger amplitudes for the diurnal period. This correlation between amplitude and pole distance (90° minus absolute latitude) cannot be seen for the sub-diurnal periods. From the error bars of the amplitudes we estimate the overall precision of the VLBI TEC values to ±5 to ±7 TECU.

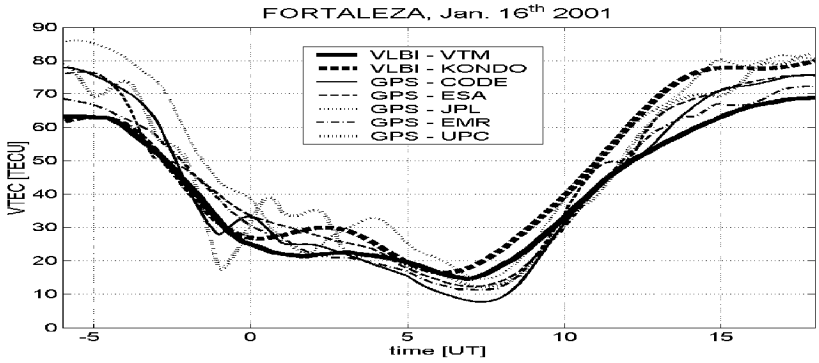


Fig. 1 VTEC over the station Fortaleza, Brasil, comparison between VLBI (Vienna TEC Model and Kondo approach) and individual GPS solutions.

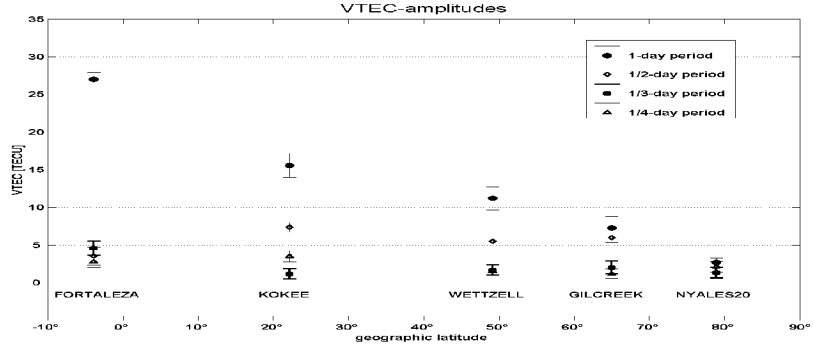


Fig. 2 Amplitudes A_{ik} and their precision from the NEOS-A session on January 16th, 2001 (Kondo approach).

The same session as before was used for deriving VTEC numbers for station Gilcreek (figure 3) which were compared to the GPS computations. In spite of the generally good agreement between the various GPS solutions and the two VLBI approaches, the residual discrepancies should be subject to a detailed investigation.

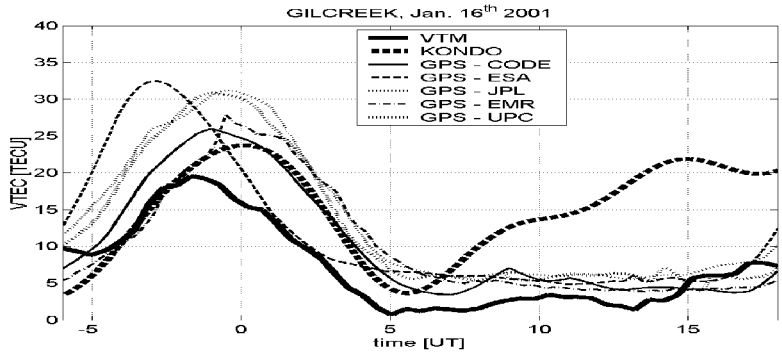


Fig. 3: VTEC over the station Gilcreek, Alaska, U.S.A., comparison between VLBI (Vienna TEC model and Kondo approach) and individual GPS solutions.

4.3 Analyzing the CONT02 session with the VTM approach

CONT02 was a VLBI session observing 15 days continuously in October 2002. The goal of the CONT02 campaign was to acquire the best possible state of the art VLBI data over a two-week period to demonstrate the highest accuracy of which VLBI is capable. The solutions of all IGS IONEX analysis centers for station Wettzell are plotted in figure 4 as dashed lines and compared with the VTM result (thick line). Figure 5 is an amplified part showing the results for station Wettzell for a period of 24 hours starting on 25th of October, 00:00 UT, 2002 in comparison with the solution derived at IGG by BERNESE software. In section 4.1 it was mentioned that so far no combined IONEX solution exists. Thus, the VLBI results could only be compared with the individual GPS solutions. For all stations in the CONT02 network the differences between the VLBI and GPS solutions were calculated every 0.1 hour and plotted in a histogram (figure 6). This should give a first idea about the accuracy of the VTM approach. Further statistical analyses will follow.

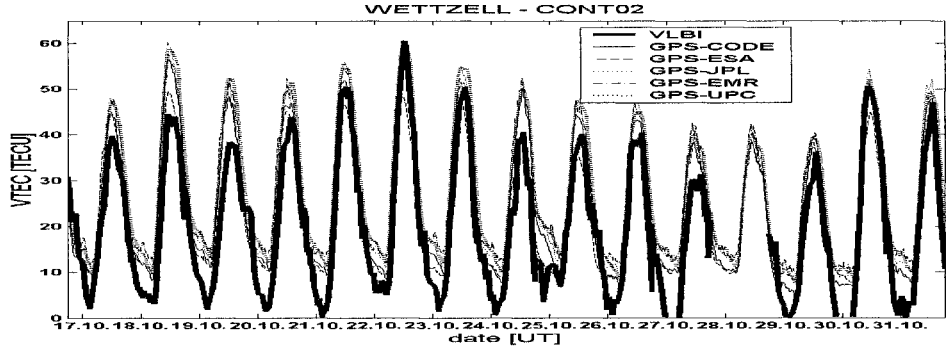


Fig. 4 VTEC over station Wettzell; comparison between VLBI (VTM approach) and GPS solutions of all IONEX analysis centers of the IGS.

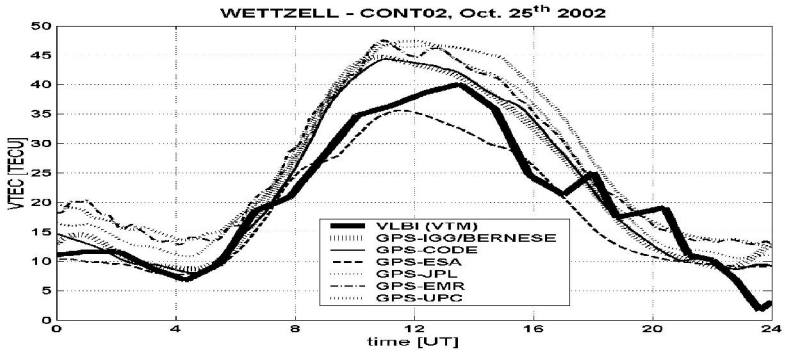


Fig. 5 VTEC over station Wettzell, comparison between VLBI (VTM approach), GPS solutions of all IONEX analysis centers and a regional TEC model derived at IGG by BERNESE GPS software .

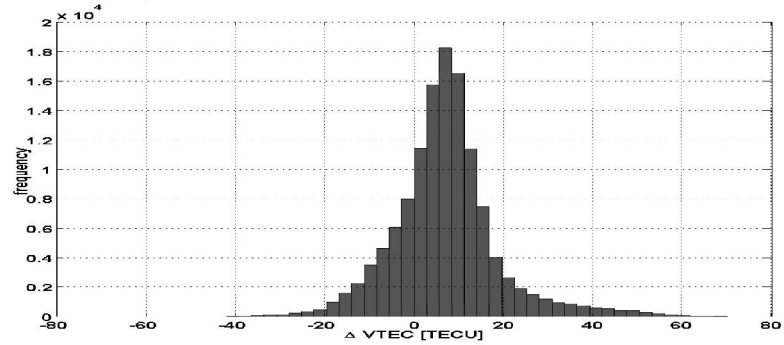


Fig. 6 Histogram of the differences between all GPS and VLBI solutions for CONT02.

The histogram shows a mean difference ('bias') between VTEC values derived by GPS and VLBI of about +7 TECU. This constant effect has to be subject to further investigations.

4 Outlook

First investigations have shown that VLBI is able to deliver information about the spatial distribution and temporal variation of the ionosphere. Further research should deal with the problem that the calculated VTEC values are not vertically located above the station. Concerning the GPS results combined VTEC series have not been published by IGS till summer 2003. However, it was decided at the last IGS governing board meeting to develop a solid procedure for combining the results of the various analysis centers. If these combined series will exist they can be used for a thorough comparison with the results obtained by VLBI. Comparison with other techniques (ionosondes, ionospheric tomography by CHAMP, Topex/Poseidon,...) will also be done.

Acknowledgements

The research project P161366-N06 "Investigation of the ionosphere by geodetic VLBI" (VLBIonos) is funded by the Austrian Science Fund (FWF).

References

Hargreaves, J.K., The solar-terrestrial environment, Cambridge atmospheric and space science series, Cambridge University Press, 1992.

Hobiger T., Boehm J., Schuh H., VLBIONOS – Probing the ionosphere by means of Very Long Baseline Interferometry, Oesterreichische Zeitschrift fuer Vermessung & Geoinformation (VGI), 1/2003, ed. by H. Schuh, F. Brunner, and H. Kahmen, 29-37, 2003.

Hartmann, G.K., Leitinger, R., Range Errors Due to Ionospheric and Tropospheric Effects for Signals above 100 MHz, Bulletin Géodesique, Nr. 58, 109-136, 1984.

Kondo T., Application of VLBI data to measurements of ionospheric total electron content, Journal of the Communications Research Laboratory, Vol .38, No. 3, 613-622, Tokyo, Japan, 1991.

Schaer, S., Gurtner, W., Feltens, J., IONEX: The Ionosphere Map EXchange Format Version 1, Proceedings of the IGS AC Workshop, Darmstadt, Germany, 9.-11. Februar, 1998, ftp://igscb.jpl.nasa.gov/igscb/data/format/ionex1.

Numerical Simulations for Geodetic Parameter Estimation from Space VLBI Data

Madhav N. Kulkarni, Praveen Pillai
Indian Institute of Technology Bombay, India, 91-22-25720456
e-mail: kulkarni@civil.iitb.ac.in

Sándor Frey
FÖMI Satellite Geodetic Observatory, Penc, Hungary, 36-27-374980
e-mail: frey@sgo.fomi.hu

Abstract. Space Very Long Baseline Interferometry (SVLBI) is primarily a radio astronomical technique, which extends conventional VLBI baselines into the space. Its potential applications in Global Geodesy include interconnection of celestial and terrestrial reference frames, and orbit determination using the delay and delay rate observables. In the past two decades, there have been significant theoretical developments in this field. With the launch of the first SVLBI satellite of the VLBI Space Observatory Programme (VSOP), Japan, this technique has now been realised. The Geodesy Demonstration Experiment (GEDEX) is designed to explore the feasibility of the geodetic applications of SVLBI. In this paper, some results of the numerical simulations for geodetic parameter estimation from VSOP data are presented.

With the availability of real data from VSOP, it is necessary to repeat the precision analysis carried out earlier, in order to re-estimate the precision under realistic conditions. Numerical simulations for this purpose have been initiated, and the first results are presented in this paper. Conclusions are drawn, based upon the results obtained, and the implications for possible dedicated geodetic SVLBI missions in the future are discussed.

Keywords. Space VLBI, GEDEX, VSOP, geodesy

1 Introduction

SVLBI is a new astronomical observing technique which extends the baselines of the traditional ground-based VLBI technique into space by means of placing a radio antenna (one of the interferometer elements) into Earth orbit. At a given observing frequency, the imaging angular resolution of the SVLBI is superior to ground-based VLBI, and presently the highest among the astronomical techniques. This enables the astronomers to study the innermost regions of distant energetic celestial radio sources (active galactic nuclei, quasars). Selected quasars form the International Celestial Reference Frame (ICRF), defined and maintained by precise position measurements using ground-based VLBI technique. The geodetic and geodynamic applications of VLBI include the determination of Earth rotation parameters (ERP's), orbital parameters, and monitoring the crustal movements.

The potential geodetic and geodynamic applications of SVLBI were pointed out by Fejes et al. (1986). The space VLBI observables, the time delay of the radio signal detected by different radio antennas, and its time derivative, the delay rate, contain information on the ground station – satellite – radio source geometry. In this respect, SVLBI is unique among the satellite geodetic techniques. It is in principle capable of directly tieing geocentric terrestrial and celestial reference frames, determining the geocentric positions of the ground VLBI stations and improving the SVLBI satellite orbit determination accuracy. A number of aspects of these potential applications were studied in detail in the early 1990's (e.g. Ádám 1990; Kulkarni 1992; Kulkarni & Ádám 1993).

The first dedicated SVLBI satellite HALCA was successfully launched from Japan under the VSOP project, in 1997 (Hirabayashi et al. 1998). The satellite was developed and is being operated by the Institute of Space and Astronautical Science, Japan, in cooperation with many organizations and radio observatories in the world. It carries an 8-m diameter antenna on board. The orbital period of the HALCA satellite is 6 hours. The observing frequencies are 1.6 and 5 GHz. Large fraction of the scientific observing time was made available to the global scientific community.

In order to explore the feasibility of the geodetic and geodynamic SVLBI applications with real data for the first time, an experiment was proposed and accepted for VSOP. The SVLBI Geodesy Demonstration Experiment (GEDEX; Kulkarni et al. 1998; Fejes 1999; Kulkarni et al. 2002) proved that, despite the practical difficulties, it is possible to obtain and analyse SVLBI delay and delay rate data. However, the observing parameters and setup of VSOP, an astrophysical space mission, were by far not optimal for geodetic analysis, and therefore the results obtained from test observations are not

accurate enough (Frey et al. 2002).

2 Problems with Real VSOP Data

There are a number of practical problems with VSOP data from the point of view of geodetic analysis (ibid). Some of these are minor obstacles (e.g. the lack of surface meteorological data at the ground stations which affects the troposphere modeling). However, several major problems also exist. As a result, the current realisation of SVLBI cannot be considered as an 'extension' to the traditional ground-based geodetic VLBI, as was often assumed in early studies (Fejes et al. 1996, and references therein). In other words, although the space baselines are indeed added to the interferometer network, on the other hand, some crucial features are missing because the VSOP mission is designed for astrophysical (radio imaging) purposes. These features are the following:

Good temporal and spatial sampling: Ground-based geodetic VLBI measurements use day-long observing runs, rapidly switching between many different radio sources covering the whole sky. On the contrary, due to the limited re-pointing capability of the satellite, VSOP observations involve one radio source. There are often gaps of up to a few hours due to the incomplete ground tracking coverage.

Bandwidth synthesis: The precision of the delay observable is dramatically increased with this technique for ground-based VLBI, but it is not available for SVLBI. The loss of accuracy due to the lack of the bandwidth synthesis and also due to the relatively small collecting area of the 8-m antenna is about two orders of magnitude with respect to an otherwise similar ground-based experiment.

Dual observing frequency: This method is used by ground-based VLBI to correct for the effect of the dispersive (ionospheric) delay. For the ground antennas in the SVLBI network, the ionospheric effects have to be modeled, or independently derived ionospheric total electron content models have to be applied (e.g. using GPS technique).

Instrumental phase calibration: While the proper separation of the instrumental and geometric delays is not crucial for most astrophysical experiments, it is important for geodetic and astrometric observations. For VSOP experiments, reliable phase-cal measurements are typically unavailable.

3 Simulations with OSUSVLBI Software using realistic input parameters

Simulation of experiments to estimate the parameters of geodetic and geodynamic interest from Space VLBI observations have been carried out (Ádám 1990; Kulkarni 1992; Kulkarni & Ádám 1993). The process of simulation analysis was divided into three distinct tasks: simulation of the orbit of Space VLBI satellite, simulation of the Space VLBI ground to space time delay observations, using the satellite orbit simulated in the first task, and adjustment computations for estimation of geodetic and geodynamic parameters from the observations simulated in the second task. Early simulation studies (Ádám 1990; Kulkarni 1992; Kulkarni & Ádám 1993) could not consider the parameters of the practical implementation of SVLBI. As it turned out later, this theoretical approach led to the overestimation of the precision achievable by the new technique. With the availability of real VSOP data (e.g. Frey et al. 2002) and in view of the realistic observing setup, we are in a position to repeat some of the simulations, and to compare them with earlier results.

The OSUSVLBI software (Kulkarni 1992) was used for the simulation studies presented here, for estimating the sensitivity to number of parameters.

The general approach followed in the numerical investigation to analyse the sensitivity of the precision of the estimated ERP's to the number of parameters like observing stations, epochs etc is same as previously done (Kulkarni 1992) but with more realistic conditions, now available. The results of these numerical tests are presented in Table 1 through Table 3 and have been plotted in Fig. 1 and Fig. 2. As explained above, only one radio source and maximum 6 observing stations are considered, based upon the realistic observing conditions.

Table 1. Effect of Number of Stations

Sr. No	No. of Stns	No. of Epochs	No. of Sources	RMS of ERP'S Xp(mas)
1	3	600	1	1.0439
	4	600	1	0.9836
	5	600	1	0.9452
	6	600	1	0.9191
2	3	900	1	0.2028
	4	900	1	0.1924
	5	900	1	0.1858
	6	900	1	0.1813
3	3	1200	1	0.0732
	4	1200	1	0.0699
	5	1200	1	0.0678
	6	1200	1	0.0663
4	3	1500	1	0.0509
	4	1500	1	0.0484
	5	1500	1	0.0469
	6	1500	1	0.0458

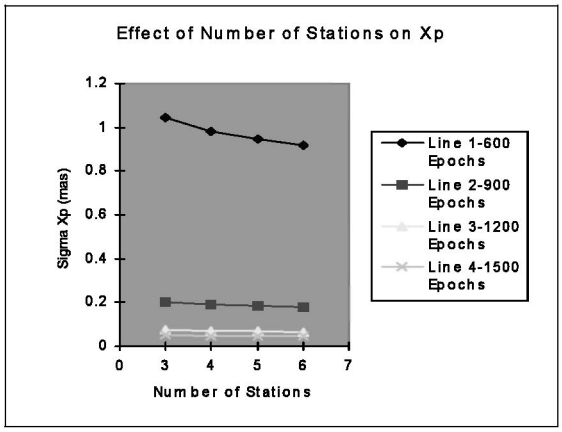


Fig. 1

Table 2. Effect of Number of Observation Epochs

Sr. No	No. of Epochs	No. of Stns.	No. of Sources	RMS of ERP'S Xp(mas)
1	600	3	1	1.0439
	900	3	1	0.2028
	1200	3	1	0.0732
	1500	3	1	0.0509
2	600	4	1	0.9836
	900	4	1	0.1924
	1200	4	1	0.0699
	1500	4	1	0.0484
3	600	5	1	0.9452
	900	5	1	0.1858
	1200	5	1	0.0678
	1500	5	1	0.0469
4	600	6	1	0.9191
	900	6	1	0.1813
	1200	6	1	0.0633
	1500	6	1	0.0458

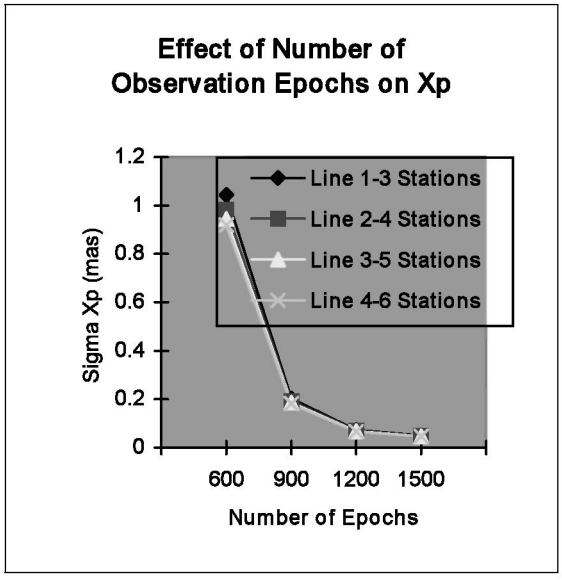


Fig. 2

Table 3. Optimum Number of Parameters Based Upon Realistic Observing Conditions

Parameter	Optimum Number
Numbar of Observation Epochs	900
Number of Observing Station	5
Number of Radio Sources	1
Total Observation Time	7.5 hours (with 30 sec observation interval, from real data)

4 Future SVLBI missions and Geodetic Data Collection Experiments

A revolutionary concept of next generation SVLBI satellite is the Satellite for Unification of Reference Frames (SURF), presented by (Fejes et al. 1996). The projected SURF applications include: orbit determination accuracy improvement, monitoring of reference frame ties (through estimation of ERPs), gravity field mapping, monitoring of geocenter positions, large angular separation phase referencing, astrometry, etc. The main requirements of SURF are: regular shape, small area-to-mass ratio, precise tracking capabilities resulting in cm-level orbit accuracy, negligible slew times and long operational life (10-15 years). The SURF design characteristics include: hexahedric or tetrahedric shape, multiple beam telescope system, phased array antennas, preselected targets, 3 axis stabilized, drag-free system, eliminating non-gravitational perturbation effects (or a set of microaccelerometers), ground antenna clusters, etc. (see Fig. 3).

Radio Astronomical Satellite (RadioAstron) is a Russian Space VLBI mission, approved in 1980s, but understandably hampered by the various problems faced by the Russian Space Agency (RKA). VLBI Space Observatory Programme (VSOP-2) is a Japanese study - a follow-on mission to the present VSOP mission, sponsored by ISAS - with launch planned for 2007. The Astronomical Low Frequency Array (ALFA) is designed to produce radio images at unexplored frequencies which are too low to be observed from the ground due to absorption in the Earth's ionosphere. It will consist of a cluster of several small satellites operating together as a single-radio telescope far from Earth. Astronomical Radio Interferometry between Space and Earth (ARISE) is a JPL Technology pre-project for a 2008 Space VLBI mission. It is based upon a 4 to 8 Gbps down-link data rate, a 25-meter inflatable antenna, and observations to 86 GHz frequency from an apogee

of up to 50,000 km. The European Space Agency (ESA) has sponsored a study of a Space VLBI mission which would rely on a 30-meter antenna assembled on-board the Space Station Freedom, to be released into a 40,000 km elliptical Earth orbit. This mission, envisioned for 2010, would include observations to 86 GHz and a down-link rate of 8 Gbps. The follow-on to RadioAstron of Russia, Millemetron, would observe in the millimeter band with a 10-meter antenna from an apogee of 300,000 km, and a data rate of 1 to 8 Gbps.

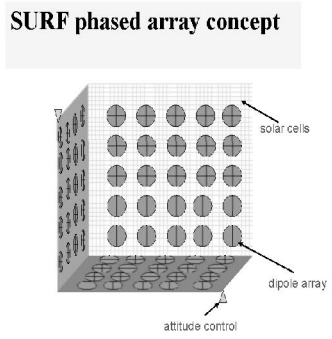


Fig. 3 Concept of a SVLBI Satellite for Unification of Reference Frames (SURF)

5 Conclusions

From these numerical investigations using more realistic parameters, conclusions can be drawn about the optimum number of different parameters like observing stations, sources, etc., for precise estimation of ERP's. The optimum total number of observation epochs would be about 900, the optimum number of observing stations, based upon realistic conditions, would be 5, and the number of radio sources being observed, from practical consideration, would be one. To summarize the results of our sensitivity analysis, the optimum numbers of parameters are tabulated in Table 3.

Acknowledgements

We gratefully acknowledge the contribution of the VSOP Project, which is led by the Institute of Space and Astronautical Science (Japan), in cooperation with many organizations and radio telescopes around the world. Financial support was received from the Hungarian Space Office, the Hungarian Scientific Research Fund (OTKA T031723) and the Indo-Hungarian Science and Technology Program. The support and funding for the research work, provided by the Department of Science & Technology, Government of India, and the Satellite Geodetic Observatory, Penc, Hungary, are gratefully acknowledged.

References

Ádám J. (1990): Estimability of Geodetic Parameters from Space VLBI Observables, Report No. 406, Dept. of Geodetic Science & Surveying, The Ohio State University, Columbus, Ohio, USA

Fejes I. (1999): The status and perspective of the Space VLBI Geodesy Demonstration Experiment (GEDEX), *Adv. Space Res.* 23, 781

Fejes I., Kawaguchi N., Mihály Sz. (1996): Space VLBI Geodesy: background of an experiment proposal, *Astrophys. Space Sci.* 239, 275

Frey S., Meyer U., Fejes I., Paragi Z., Charlot P., Biancale R. (2002): Geodetic space VLBI: the first test observations, *Adv. Space Res.* 30, 357

Hirabayashi H., Hirosawa H., Kobayashi H., et al. (1998): Overview and Initial Results of the Very Long Baseline Interferometry Space Observatory Programmed, *Science* 281, 1825

Kulkarni M.N. (1992): A Feasibility Study of Space VLBI for Geodesy & Geodynamics, Report No. 420, Dept. of Geodetic Science & Surveying, The Ohio State University, Columbus, Ohio, USA

Kulkarni M.N., Ádám J. (1993): A Sensitivity Analysis for Orbit Determination of SVLBI Satellites for Interconnection of Reference Frames, *Acta Geod. Geophys. Mont. Hung.* 28, 441

Kulkarni M.N., Ádám J., Fejes I., Frey S., Kampes B., Zheng Y., Hu X. (1998): Geodesy Demonstration Experiment (GEDEX) for Space VLBI: State of the Art and software Development. in: Geodesy on the Move - Gravity, Geoid, Geodynamics and Antarctica, eds. Forsberg R., Feissel M., Dietrich R., *IAG Symposia* 119, Springer-Verlag, Berlin-Heidelberg-New York, 383

Kulkarni M.N., Ádám J., Fejes I., Frey S., Paragi Z. (2002): Recent Developments in Space VLBI Geodesy Experiment (GEDEX), in: Vistas for Geodesy in the New Millennium, eds. Ádám J., Schwarz K.-P., *IAG Symposia* 125, Springer-Verlag, Berlin-Heidelberg-New York, 18

Relativistic VLBI Delay Model for Finite Distance Radio Source

Mamoru Sekido[1] and Toshio Fukushima[2]

1: Communications Research Laboratory, 893-1 Hirai Kashima Ibaraki, 314-0012 Japan, e-mail: sekido@crl.go.jp
2: National Astronomical Observatory, 2-21-1 Osawa Mitaka Tokyo, 181-8588 Japan, e-mail: Toshio.Fukushima@nao.ac.jp

Abstract. When very long baseline interferometry (VLBI) is applied to observation of finite distance radio source from observer and the curvature of the the wave front cannot be approximated as plane wave. Thus current standard VLBI model ('Consensus model') does not have enough accuracy and an alternative precise VLBI delay model for finite distance radio source corresponding to the 'Consensus model' is required in observation of radio source in the solar system (e.g. planets, asteroids, and spacecraft). We derived a formular of relativistic VLBI delay model for finite distance radio source by taking into account coordinates transformation by using linearized PPN metric. This model is valid in order of several pico seconds accuracy when the radio source is at distance beyond 10^9 m. Our model include the 'Consensus model' as a special case that radio source is at infinite distance form the observer.

Key words. Very Long Baseline Interferometry – Theory of Relativity – Finite Distance – Spacecraft Navigation

1 Introduction

Accurate delay model is required in VLBI data processing and analysis, especially for precise astrometry and geodesy. Based on plane wave approximation, VLBI observation models including relativistic effects are proposed from 1980s to beginning of 1990s. Finally, T. M. Eubanks has summarized those models and proposed a 'consensus model'(Eubanks, 1991) in 1991. The 'Consensus model' is widely used in VLBI community in the world as a standard VLBI model (McCarthy and Petit, 2003, chap. 11)[1].

However, this standard VLBI model does not have enough accuracy, when radio source is at the distance no further than 30 light years (e.g. planets, asteroids, and spacecraft in the solar system), or in observation of relatively closer radio source (a hundred light year away) with space VLBI (Ground - satellite or stellite - satellite) baseline. Because curvature of the wave front cannot be ignored in those observations, then alternative VLBI delay model is required for finite distance radio sources.

Sovers and Jacobs (1996) discussed on curvature effect of finite distance radio source. Fukushima (1994) proposed useful expression of VLBI delay model for finite distance radio source and discussed on partial derivatives. However, analytical formula of relativistic VLBI delay model corresponding to the 'Consensus model' (Eubanks, 1991; McCarthy and Petit, 2003, chap. 11) was not on their papers. Moyer (2000) has developed formulation of radiometric observation data for spacecraft navigation with based on light time equation. Solving light time equation by numerical procedure is straightforward and precise enough, however it need iteration of computation to solve light time equation in the analysis software. Our approach is aimed to derive a formula corresponding to the 'Consensus Model' to be implemented in CALC[2] program without major modification of the software, even some accuracy is lost as compromise. We followed approach of Hellings (1986), Shahid-Saless and Hellings (1991), and Fukushima (1994) for the derivation of an analytical formula in this paper. Our target is order of pico seconds accuracy with ground based baseline for radio source in the solar system.

2 Delay in the rest frame of Solar System Barycenter

Variables of large capital indicate quantity in the rest frame of solar system barycenter (hereafter referred as FCB) and small ones represent those of geocentric reference frame (hereafter referred as FCG). Then Barycentric Coordinate Time (TCB) is represented by T and Geocentric Coordinate Time (TCG) is represented by t.

[1] http://maia.usno.navy.mil/conv2000.html

[2] http://gemini.gsfc.nasa.gov/solve/solve.shtml

Suffix 0,1,2 represent radio source and two VLBI observation stations, respectively. Position vector of i^{th} station in the FCB is expressed by $\mathbf{X}_i$. Radio signal is supposed to be emitted at T_0 from the radio source and to arrive at i^{th} observation station at T_i.

$$\mathbf{R}_{ij} = \boldsymbol{X}_i - \boldsymbol{X}_j, \quad R_{ij} = |\mathbf{R}_{ij}|$$

Let us define a pseudo baseline vector $\mathbf{B}^*$ and pseudo source vector $\mathbf{K}^*$ as follows:

$$\begin{aligned} \mathbf{B}^* &= \boldsymbol{X}_2(T_2) - \boldsymbol{X}_1(T_1) = \boldsymbol{R}_{20}(T_2) - \boldsymbol{R}_{10}(T_1) \\ \mathbf{K}^* &= \frac{\mathbf{R}_{02}(T_2) + \mathbf{R}_{01}(T_1)}{R_{01}(T_1) + R_{02}(T_2)}. \end{aligned} \tag{1}$$

Then time interval between arrival of signal to two stations is expressed as follows (Fukushima, 1994):

$$\begin{aligned} c(T_2 - T_1) &= R_{02}(T_2) - R_{01}(T_1) + c\Delta t_g \\ &= -\mathbf{K}^* \cdot \mathbf{B}^* + c\Delta t_g, \end{aligned} \tag{2}$$

where Δt_g indicates difference of delay due to gravitational refraction of light between two VLBI stations. Predicted coordinates of radio source is supposed to be given apriori as function of time. The coordinates of radio source $\boldsymbol{X}_0(T_0)$, where a signal is departed at T_0 and arrived to station $\boldsymbol{X}_1$ at T_1, is obtained by a solution of light time equation

$$\begin{aligned} c(T_1 - T_0) &= |\mathbf{X}_1(T_1) - \mathbf{X}_0(T_0)| \\ &+ (1+\gamma) \sum_J \frac{GM_J}{c^2} \ln \left| \frac{R_{0J} + R_{1J} + R_{01}}{R_{0J} + R_{1J} - R_{01}} \right| \end{aligned} \tag{3}$$

Let baseline vector and pseudo source vector defined at the same epoch T_1 in TCB be respectively, $\boldsymbol{B}$ and $\boldsymbol{K}$. We want to express the delay time of equation (2) with $\boldsymbol{B}$ and $\boldsymbol{K}$. Since the time interval $T_2 - T_1$ is small (43 msec for earth diameter), $\mathbf{X}_2(T_1)$ can be expressed by eliminating higher order terms than velocity as:

$$\boldsymbol{X}_2(T_2) = \mathbf{V}_2(T_2 - T_1) + \boldsymbol{X}_2(T_1) \tag{4}$$

$$\mathbf{R}_{02}(T_2) = \mathbf{R}_{02}(T_1) - \mathbf{V}_2(T_2 - T_1) \tag{5}$$

$$\begin{aligned} R_{02} &= |\mathbf{R}_{02}(T_1) - \mathbf{V}_2(T_2 - T_1)| \\ &\cong R_{02}(T_1) - \widehat{\mathbf{R}}_{02} \cdot \mathbf{V}_2(T_2 - T_1) \\ &+ \frac{[V_2(T_2 - T_1)]^2 - [\widehat{\mathbf{R}}_{02} \cdot \mathbf{V}_2(T_2 - T_1)]^2}{2R_{02}(T_1)} \end{aligned} \tag{6}$$

where, $\widehat{\mathbf{R}}_{ij} = \mathbf{R}_{ij}/R_{ij}$. Square root of the equation was approximated up to second order of $V_2(T_2 - T_1)/R_{02}$. In ground based VLBI observation of radio source at distance beyond 10^9 m, the contribution from the second order term is less than 1.5 mm. Thus we will keep remain up to the second term of equation (6). Substituting this into equation (2) becomes

$$\begin{aligned} &c(T_2 - T_1) \\ &= R_{02}(T_1) - R_{01}(T_1) - \widehat{\mathbf{R}}_{02} \cdot \mathbf{V}_2(T_2 - T_1) + c\Delta t_g \\ &= -\mathbf{K} \cdot \mathbf{B} - \widehat{\mathbf{R}}_{02} \cdot \mathbf{V}_2(T_2 - T_1) + c\Delta t_g. \end{aligned}$$

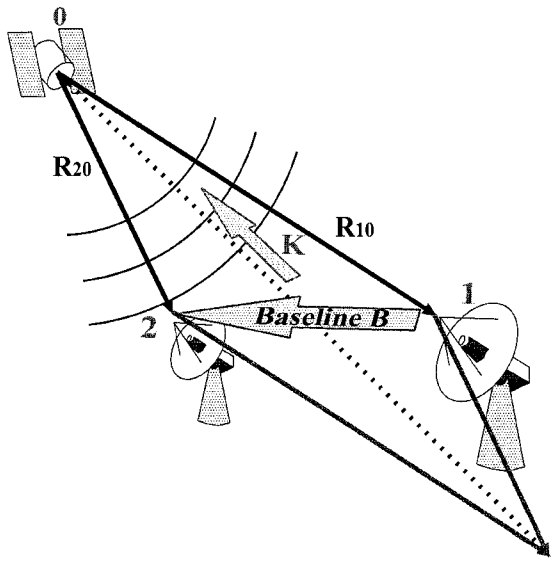


Fig. 1. Schematic diagram of pseudo source vector **K**. The vector **K** is on the diagonal line of parallelogram composed from vector $\mathbf{R}_{10}$ and $\mathbf{R}_{20}$.

(7)

Therefore, VLBI delay for finite distance radio source expressed in TCB is

$$T_2 - T_1 = \frac{1}{1 + \beta_{02}} \left(-\frac{\mathbf{K} \cdot \mathbf{B}}{c} + \Delta t_g \right) \tag{8}$$

where,

$$\beta_{02} = \widehat{\mathbf{R}}_{02} \cdot \frac{\mathbf{V}_2}{c} \tag{9}$$

$$\mathbf{B} = \boldsymbol{X}_2(T_1) - \boldsymbol{X}_1(T_1) \tag{10}$$

$$\mathbf{K} = \frac{\mathbf{R}_{02}(T_1) + \mathbf{R}_{01}(T_1)}{R_{01}(T_1) + R_{02}(T_1)} \tag{11}$$

An attention have to be paid on the vector **K**, which should be called as pseudo source vector because it is neither unit vector nor constant vector as the ordinary source vector is. The geometrical image of this vector is sketched in figure 1. The direction of **K** vector is on the diagonal line of parallelogram composed from two vectors $\mathbf{R}_{01}$ and $\mathbf{R}_{02}$, thus the direction and amplitude of the pseudo source vector depend on time and baseline. Behavior of **K** vector and difference from ordinary source vector is discussed in section 4.

3 Transformation from Barycentric Coordinates to Geocentric Coordinates

General relativity tells that coordinate transformation in gravity field belongs to a group of general coordinate transformation, which keep infinitesimal line element of events $(ds)^2 = g_{\mu\nu} dx^\mu dx^\nu$ invariant. Here we

use a metric tensor $g_{\mu\nu}$ of post-Galilean (linearized post-Newtonian) approximation by a following approach of Hellings(1986)

$$\begin{aligned} g_{00} &= 1 - 2\phi + O(c^{-4}) \\ g_{0k} &= O(c^{-3}) \\ g_{mn} &= -\delta_{mn}(1 + 2\gamma\phi) + O(c^{-4}), \end{aligned} \tag{12}$$

where $\phi = \sum_p \frac{GM_p}{|\mathbf{X}-\mathbf{X}_p|c^2}$ is summation of gravitational potential at $\mathbf{X}$ produced by p^{th} body. γ is one of Parameterized Post Newtonian parameters, and is unity in general relativity. Higher order terms than c^{-2} are eliminated, although it is enough for our target accuracy.

Infinitesimal coordinates transformation between FCB $(T, \mathbf{X})$ and a FCG coordinates nearby earth $(t, \mathbf{x})$ is given by Hellings (1986) as

$$\begin{aligned} dt &= (1 - U + \frac{V_e^2}{2c^2})dT - (1 + \gamma U + \frac{V_e^2}{2c^2})\frac{\mathbf{V}_e \cdot d\mathbf{X}}{c^2} \\ d\mathbf{x} &= (1 + \gamma U)(d\mathbf{X} + \frac{d\mathbf{X} \cdot \mathbf{V}_e}{2c^2}\mathbf{V}_e) - (1 - U + \frac{V_e^2}{2c^2})\mathbf{V}_e dT. \end{aligned} \tag{13}$$

And its inverse transformation is

$$\begin{aligned} dT &= (1 + U + \frac{V_e^2}{2c^2})dt + (1 - \gamma U + \frac{V_e^2}{2c^2})\frac{\mathbf{V}_e \cdot d\mathbf{x}}{c^2} \\ d\mathbf{X} &= (1 - \gamma U)(d\mathbf{x} + \frac{d\mathbf{x} \cdot \mathbf{V}_e}{2c^2}\mathbf{V}_e) + (1 + U + \frac{V_e^2}{2c^2})\mathbf{V}_e dt. \end{aligned} \tag{14}$$

Baseline vector of equation (10) is defined in simultaneity in TCB ($dT = 0$), however, this is not simultaneous in TCG. Corresponding events are $(T_1, \mathbf{X}_1) \Leftrightarrow (t_1, \mathbf{x}_1)$ and $(T_1, \mathbf{X}_2) \Leftrightarrow (t_1^*, \mathbf{X}_2)$. Time difference $t_1^* - t_1$ is derived from the first equation of (13) with condition $dT = 0$ as

$$\begin{aligned} t_1^* - t_1 &= -(1 + \gamma U + \frac{V_e^2}{2c^2})\frac{\mathbf{V}_e \cdot \mathbf{b}}{c^2} \\ &\cong -\frac{\mathbf{V}_e \cdot \mathbf{b}}{c^2} \end{aligned} \tag{15}$$

By integrating the second equation of (14) and by using the relation (15), baseline vector defined by equation (10) is expressed by

$$\begin{aligned} \mathbf{B} &= \mathbf{X}_2(T_1) - \mathbf{X}_1(T_1) \\ &= (1 - \gamma U)(\mathbf{b} - \frac{\mathbf{V}_e \cdot \mathbf{b}}{c^2}\mathbf{w}_2) \\ &\quad + (1 - \gamma U)\frac{[\mathbf{b} - \frac{\mathbf{V}_e \cdot \mathbf{b}}{c^2}\mathbf{w}_2] \cdot \mathbf{V}_e}{2c^2}\mathbf{V}_e \\ &\quad - (1 + U + \frac{V_e^2}{2c^2})\frac{\mathbf{V}_e \cdot \mathbf{b}}{c^2}\mathbf{V}_e \\ &\cong (1 - \gamma U)\mathbf{b} - \frac{\mathbf{V}_e \cdot \mathbf{b}}{c^2}(\frac{\mathbf{V}_e}{2} + \mathbf{w}_2) \end{aligned} \tag{16}$$

Relation of time interval between $T_2 - T_1$ and $t_2 - t_1$ is derived form the first equation of (14) as follows:

$$\begin{aligned} T_2 - T_1 &= (1 + U + \frac{V_e^2}{2c^2})(t_2 - t_1) \\ &\quad + (1 + \gamma U + \frac{V_e^2}{2c^2})\frac{\mathbf{V}_e \cdot [\mathbf{x}_2(t_2) - \mathbf{x}_1(t_1)]}{c^2} \\ &= (1 + U + \frac{V_e^2}{2c^2} + \frac{\mathbf{V}_e \cdot \mathbf{w}_2}{c^2})(t_2 - t_1) \\ &\quad + (1 + \gamma U + \frac{V_e^2}{2c^2})\frac{\mathbf{V}_e \cdot \mathbf{b}}{c^2}, \end{aligned} \tag{17}$$

where, a relation $\mathbf{x}_2(t_2) = \mathbf{x}_2(t_1) + \mathbf{w}_2(t_2 - t_1)$ was used.

Substituting expression of baseline vector (equation (16)) and relation of time interval (equation (17)) to equation (8), we get relation of time interval $t_2 - t_1$ and baseline vector $\mathbf{b}$ as

$$\begin{aligned} (t_2 - t_1) &= (1 + \beta_{02})^{-1} \\ &\quad \left\{ \Delta t_g - \frac{\mathbf{K} \cdot \mathbf{b}}{c}\left[1 - (1 + \gamma)U - \frac{V_e^2 + 2\mathbf{V}_e \cdot \mathbf{w}_2}{2c^2}\right] \right. \\ &\quad \left. - \frac{\mathbf{V}_e \cdot \mathbf{b}}{c^2}\left(1 + \beta_{02} - \frac{\mathbf{K} \cdot (\mathbf{V}_e + 2\mathbf{w}_2)}{2c}\right)\right\} \end{aligned} \tag{18}$$

This formula relate quantity t and $\mathbf{x}$, which are quantities in geocentric reference frame. Time interval of signal arrival to VLBI stations is, however, based on time scale of terrestrial time (TT), which is defined as idealized atomic time scale on the geoid (Seidelmann and Fukushima, 1992). The clock rate of TT is slightly difference from TCG and they are related with $TT = (1 - L_G)TCG$, where $L_G = 6.969290134 \times 10^{-10}$ (McCarthy and Petit, 2003, chap. 1). By using time scale conversion from TCG (t) to TT (τ): $\tau = (1 - L_G)t$ and corresponding spatial scale conversion: $\boldsymbol{\xi} = (1 - L_G)\mathbf{x}$ to keep speed of light in constant, VLBI delay measured by clock on the geoid is found to be

$$\begin{aligned} \tau_2 - \tau_1 &= (1 + \beta_{02})^{-1} \\ &\quad \left\{ \Delta t_g - \frac{\mathbf{K} \cdot \mathbf{b}}{c}\left[1 - (1 + \gamma)U - \frac{V_e^2 + 2\mathbf{V}_e \cdot \mathbf{w}_2}{2c^2}\right] \right. \\ &\quad \left. - \frac{\mathbf{V}_e \cdot \mathbf{b}}{c^2}\left(1 + \beta_{02} - \frac{\mathbf{K} \cdot (\mathbf{V}_e + 2\mathbf{w}_2)}{2c}\right)\right\}, \end{aligned} \tag{19}$$

where baseline vector on the geoid $\mathbf{b} = \boldsymbol{\xi}_2(\tau_1) - \boldsymbol{\xi}_1(\tau_2)$ was re-defined. Used variables are

$$\begin{aligned} \beta_{02} &= \widehat{\mathbf{R}}_{02} \cdot \frac{\mathbf{V}_2}{c} \\ \widehat{\mathbf{R}}_{02} &= \frac{\mathbf{R}_{02}}{R_{02}} \\ \mathbf{K} &= \frac{\mathbf{R}_{02}(T_1) + \mathbf{R}_{01}(T_1)}{R_{02}(T_1) + R_{01}(T_1)}. \end{aligned}$$

and gravitational delay Δt_g is expressed

$$\Delta t_g = \sum_J \frac{GM_J}{c^3} \ln\left(\frac{R_{2J} - \widehat{\mathbf{R}}_{02} \cdot \mathbf{R}_{2J}}{R_{0J} - \widehat{\mathbf{R}}_{02} \cdot \mathbf{R}_{0J}} \cdot \frac{R_{0J} - \widehat{\mathbf{R}}_{01} \cdot \mathbf{R}_{0J}}{R_{1J} - \widehat{\mathbf{R}}_{01} \cdot \mathbf{R}_{1J}}\right)$$

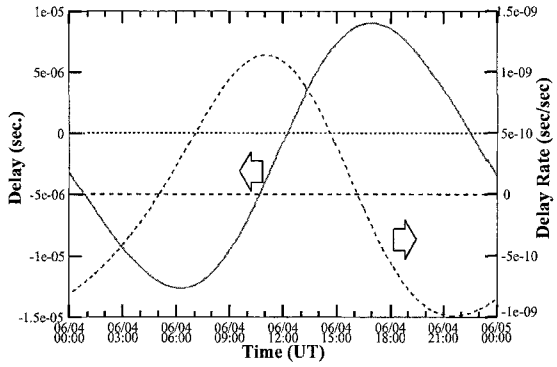


Fig. 2. Difference of delay and rate of finite VLBI delay model from ordinary VLBI delay model when spacecraft NOZOMI is observed with Kashima-Algonquin (9000 km) baseline. Consensus delay and rate model were already subtracted, where geocentric unit vector to the radio source was used as source vector. Solid line indicates delay scaled with left vertical axis. Dashed line is delay rate scaled with right vertical axis. Geocentric distance to the NOZOMI was $4. \times 10^9$ m at this time.

(20)

where, G is Newtonian gravitational constant, M_J is mass of the gravitation source J and $\mathbf{R}_{ij} = \mathbf{X}_i - \mathbf{X}_j$, $R_{ij} = |\mathbf{R}_{ij}|$, and $\widehat{\mathbf{R}}_{ij} = \mathbf{R}_{ij}/R_{ij}$. The each position epoch of gravity sources must be chosen when distance become minimum between that object and the photon (radio wave) to be measured.

4 Comparison with the 'Consensus Model'

The 'Consensus Model' (Eubanks, 1991) is

$$\tau_2 - \tau_1 = \left[1 + \frac{\mathbf{K}_s \cdot (\mathbf{V}_e + \mathbf{w}_2)}{c}\right]^{-1} \left\{\Delta t_g - \frac{\mathbf{K}_s \cdot \mathbf{b}}{c}\left[1 - (1+\gamma)U - \frac{V_e^2 + 2\mathbf{V}_e \cdot \mathbf{w}_2}{2c^2}\right] - \frac{\mathbf{V}_e \cdot \mathbf{b}}{c^2}\left(1 + \frac{\mathbf{V}_e \cdot \mathbf{K}_s}{2c}\right)\right\}, \tag{21}$$

where $\mathbf{K}_s$ is unit source vector defined in rest frame of solar system barycenter.

Finite VLBI delay model (equation (19)) and consensus model is quite similar in form. Difference from normal VLBI delay model is, however, significant as displayed in Figure 2. The figure shows difference of delay and rate between finite VLBI model and consensus model when spacecraft NOZOMI[3] is observed with Kashima-Algonquin baseline, as an example. Attention have to be paid that delay rate deviates in the order of 10^{-9}(sec/sec) from normal VLBI model, which makes it difficult to detect fringe of the interferometer without use of finite VLBI model.

[3] Spacecraft NOZOMI was launched for Mars exploration by Institute of Space and Astronautical Sciences of Japan.

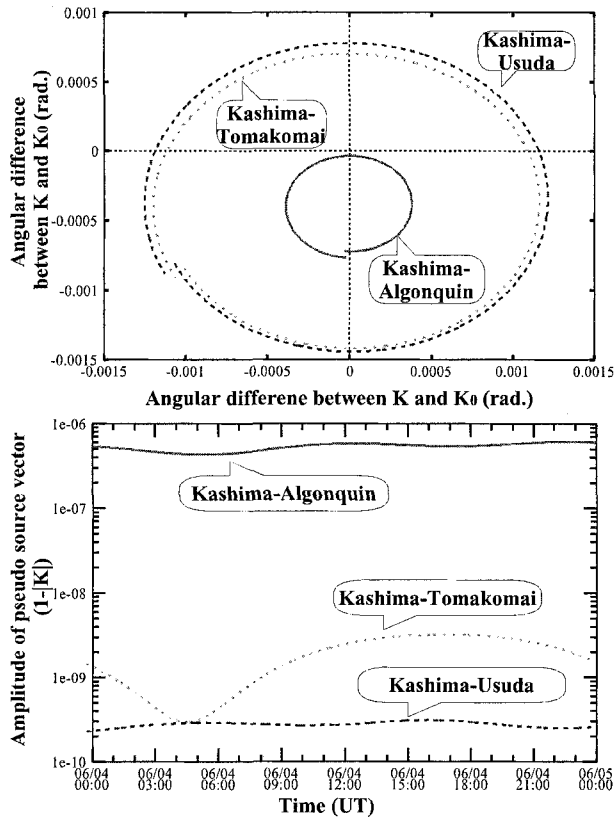


Fig. 3. Behavior of pseudo source vector $\mathbf{K}$ in comparison with geocentric unit source vector $\mathbf{K}_0$. Upper panel shows changes of the direction of $\mathbf{K}$ vector during 24 hours. The origin of the plot is the direction of $\mathbf{K}_0$ vector. Lower panel shows deviation of the magnitude of $\mathbf{K}$ vector from unity. Three lines on both two panels represent Kashima-Algonquin (9109 km), Kashima-Tomakomai (750 km), and Kashima-Usuda (208 km) baselines. Predicted orbit of NOZOMI on 4th-5th June 2003 was used for computation of the vectors here. Geocentric distance to the NOZOMI was $4. \times 10^9$ m in this period.

Unique characteristic of delay behavior of the finite VLBI delay model is mainly originated from the properties of pseudo source vector $\mathbf{K}$. Differently from ordinary source vector $\mathbf{K}_s$, the pseudo source vector $\mathbf{K}$ changes its direction and magnitude with time and baseline. Examples of behavior of $\mathbf{K}$ vector is displayed in Figure 3 in comparison with geocentric unit vector to the source $\mathbf{K}_0 \equiv \mathbf{R}_{0g}/R_{0g}$. The upper panel shows that the direction of the $\mathbf{K}$ deviates from the direction of $\mathbf{K}_0$ more significantly as baseline is shorter. It is because parallax affect more directly to $\mathbf{K}$ vector on shorter baseline. Also the NOZOMI is moving, so the track of the vectors are not closed curves in 24 hours. Magnitude of the vector $\mathbf{K}$ (lower panel) differs from unity more remarkably as baseline becomes longer. Since the (pseudo) source vector affects directly to VLBI delay, change of its direction and magnitude cause change of delay in the same order.

As it is obvious from the definition, the pseudo source vector $\mathbf{K}$ (equation (11)) and unit vector $\widehat{\mathbf{R}}_{02}$ approach to ordinary source vector $\mathbf{K}_s$ as distance to the radio source increase.

$$\lim_{R_{01},R_{02}\to\infty} \mathbf{K}(R_{01}, R_{02}) = \mathbf{K}_s \tag{22}$$

$$\lim_{R_{02}\to\infty} \widehat{\mathbf{R}}_{02} = \mathbf{K}_s \tag{23}$$

By using $\mathbf{V}_2 \cong \mathbf{V}_e + \mathbf{w}_2$ and equation (23), we can find

$$\lim_{R_{02}\to\infty} \beta_{02} = \mathbf{K}_s \cdot \frac{\mathbf{V}_2}{c} \cong \mathbf{K}_s \cdot \frac{\mathbf{V}_e + \mathbf{w}_2}{c}. \tag{24}$$

Then the last term of equation (19) approach to that of equation (21) at the limit of infinite distance ($R_{i0} \to \infty$). Therefore the finite distance VLBI model include the 'consensus model' as a special case that the distance to the radio source is infinity.

This finite VLBI model has accuracy of several pico seconds in the region that distance from observer to radio source is over 10^9 m. In case that the distance to the radio source is closer than 10^9 m, solving light time equation (3) for two paths ($0 \to 1$ and $0 \to 2$) and taking difference between them (light-time approach) might give direct way to calculate delay and delay rate precisely. Finite distance VLBI model (VLBI-like approach) will have advantage to the light-time approach in the region that distance to radio source is distant ($\geq 10^9$). Because most part of flight time of light between observation stations will canceled out as the distance increase, and consequently numerical precision might be lost in case of light-time approach.

5 Summary

Analytical formula of relativistic VLBI delay model for finite distance radio source is derived based on linearized post Newtonian metric. This formula has precision of less than five pico seconds for the radio source beyond 10^9 m from observers.

Acknowledgements. Some part of this study has performed during author's stay in (Space Geodynamics Laboratory) York University in Canada by means of grant of Japanese Ministry of Education, Culture, Sports, Science, and Technology. We thanks to W. Cannon at SGL for supporting this work. We thanks to M. Yoshikawa at ISAS and delta-VLBI team member of CRL, ISAS, NAO, Gifu Univ., Yamaguchi Univ., and Hokkaido Univ. for encouragements.

References

Eubanks, T. M. (1991) A Consensus Model for Relativistic Effects in Geodetic VLBI. Proc. of the USNO workshop on Relativistic Models for Use in Space Geodesy: 60-82

Fukushima, T. (1994) Lunar VLBI observation model. A&A 291: 320-323

Hellings, R. W. (1986) RELATIVISTIC EFFECTS IN ASTRONOMICAL TIMING MEASUREMENTS. AJ, 91: 650-659

McCarthy, D. D. and Petit, G. (2003), IERS Conventions.

Moyer, T. D. (2000) Formulation for Observed and Computed Values of Deep Space Network Data Types for Navigation, JPL Monograph 2 (JPL Publication 00-7).

Seidelmann, P. K. and Fukushima, T. (1992) Why new time scale?. A&A 265: 833-838

Sovers, O. J. & Jacobs C. S. (1996) Observation Model and Parameter Partials for the JPL VLBI Parameter Estimation Software "MODEST"-1996". JPL Publication 83-39, Rev. 6: 6–8

Shahid-Saless, B. & Hellings, R. W. (1991) A Picosecond Accuracy Relativistic VLBI Model via Fermi Normal Coordinates. GRL 18: 1139-1142

Synergy of VLBI and GPS observations in Japan

Hiromichi Tsuji, Tadashi Tanabe, Hiroshi Kawawa, Kohhei Miyagawa, Kazuhiro Takashima, Shinobu Kurihara, Yoshihiro Fukuzaki, and Shigeru Matsuzaka

Geographical Survey Institute, Kitasato-1, Tsukuba, Ibaraki, 305-0811 Japan
e-mail: tsuji@gsi.go.jp; Tel: +81-29-864-4796; Fax: +81-29-864-1802

Abstract. Although Very Long Baseline Interferometer (VLBI) operations require considerable resources compared to GPS, VLBI is an essential tool for geodesy in maintaining International Terrestrial Reference Frame and providing Earth Orientation Parameters for precise GPS surveying. This paper reviews GSI's VLBI programs focusing on the synergy with GPS in Japan. One of the largest fruits is the realization of Geodetic Coordinates 2000 based on the inter-technique solution of VLBI and GPS for national geodetic application.

Keywords: VLBI, GPS, ITRF, EOP

1 Introduction

The Geographical Survey Institute (GSI) operates four Very Long Baseline Interferometer (VLBI) stations and a correlation center in Japan for geodetic applications under International VLBI service (IVS). Table 1 and Fig. 1 show GSI's geodetic VLBI facilities.

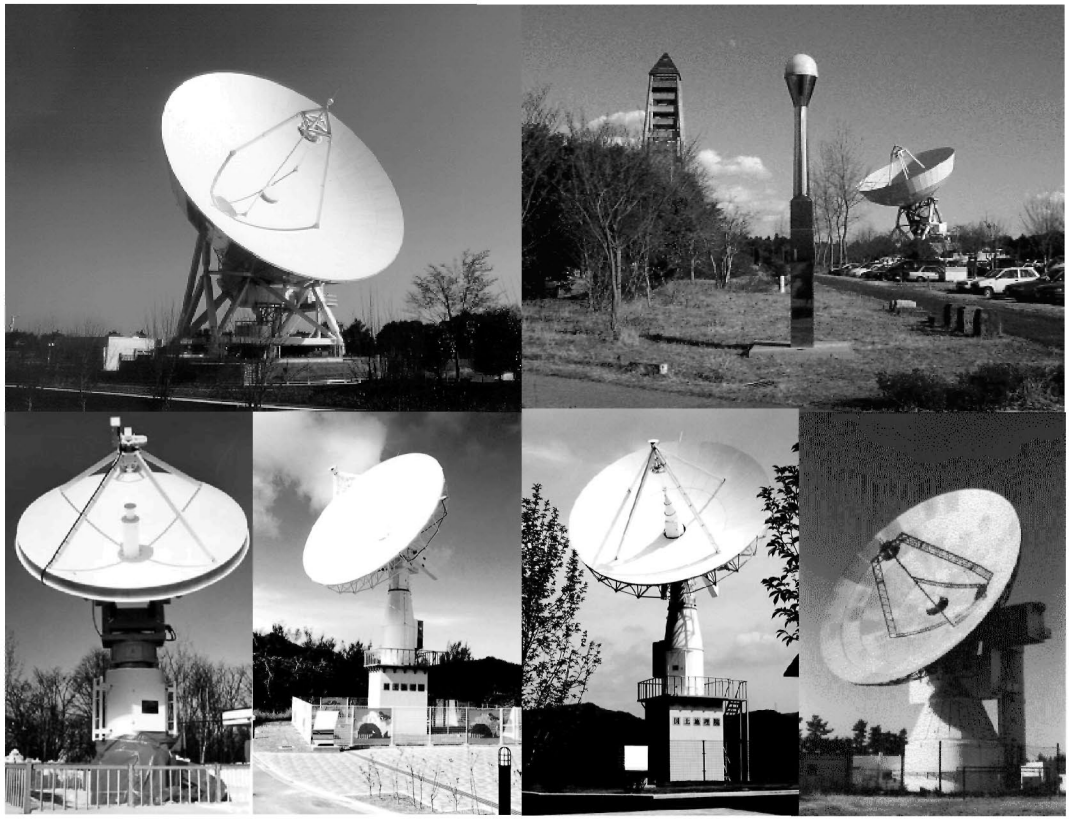


Fig. 1 The Lineup of GSI's VLBI antennas devoted to Geodesy in Japan. From top left to right: Tsukuba 32m antenna, the same antenna snapped with a GPS-based control point, Shintotsukawa 3.8m antenna, Chichijima 10m antenna, Aira 10m antenna and Kashima 26m antenna. Note that the Kashima 26m antenna retired from its 34 year service in March 2003.

Table 1. Geodetic VLBI facilities in Japan

Station	Specification	Status
Shintotsukawa, Hokkaido	3.8m diameter, S/X/K band[1], K-4 backend system	Operational from 1995
Aira, Kagoshima	10m diameter, S/X/K band, K-4 backend system	Operational from 1997
Chichijima, Ogasawara islands	10m diameter, S/X/K band, K-4 backend system	Operational from 1997
Tsukuba, Ibaraki	32m diameter, S/X/K band, Mark IV and K-4 backend systems	Operational from 1998
Tsukuba VLBI center	K-4 correlator for 3 stations, Analysis workstations, Remote control unit	Operational from 1998

[1] S-band: 2.20-2.50GHz, X-band: 7.78-8.98GHz, K-band: 19.5-25.0GHz (No backend for K-band)

2 Why VLBI in Japan ?

A wealth of earthquakes and natural hazards in Japan led GSI to the formation of a nationwide continuous GPS array in mid 1990s as a sensor to monitor crustal deformations. The array, known as GEONET (GPS Earth Observation Network), currently consists of about 1,200 GPS-based control points distributed all over Japan, providing daily site coordinates of few cm precision (Hatanaka 2002). If so, are there any roles left for VLBI in Japan ?

Actually we have a long list of answers to this important question, but it can be summarized in one phrase that GPS surveying depends on VLBI products. To be more specific, GSI dedicates 7 staff members and an operational cost of about 100 million JPY each year to VLBI because

- a set of coordinates of domestic VLBI stations defines the framework of a new geodetic datum of Japan,
- VLBI is one of few space geodetic techniques that has precision to calibrate coordinates of GEONET along with Satellite Laser Ranging (SLR),
- VLBI is the only technique that can provide a full set of Earth Orientation Parameters (EOP), that are indispensable for satellite orbit determinations including GPS at International GPS Service (IGS), and
- VLBI provides a very stable terrestrial reference frame with natural signal sources that will last over astronomical years; such a reference frame is necessary for monitoring sea level change due to global warming.

3 Programs under IVS

To achieve missions stated above, GSI actively joins and supports the IVS-based international and domestic VLBI programs such as IVS-T, IVS-R, and JADE. Table 2 summarizes IVS observation programs that GSI plans to participate in 2003. Total number of session is 72, showing a slight increase from 62 in 2002.

Table 2. IVS observation programs with GSI in 2003

Code Name	Purpose	Sessions per year[1]	Duration of session	Number of stations	Correlator
IVS-T	ITRF maintenance	4 / 12	24 hour	8	Bonn Univ.
IVS-R	EOP determination	18 / 113	24 hour	7	Bonn Univ., USNO[2], MIT
APSG	ITRF maintenance in Asia and the Pacific	2 / 2	24 hour	6	USNO
VLBA	Astrometry	6 / 6	24 hour	20	NRAO[3]
UT1-Intensive	EOP determination (UT1 only)	31 / 31	1 hour	2	GSI
JADE	ITRF maintenance in Japan	11 / 11	24 hour	4	GSI

[1] Number of sessions that GSI participates / Number of total sessions
[2] US Naval Observatory
[3] National Radio Astronomy Observatory

4 Topics

4.1 Achievements of outgoing Kashima 26m antenna

The Kashima 26m antenna, constructed in 1968 at Communication Research Laboratory (CRL) Kashima, had played a key role in Research and Development of radio communication technology and its application to VLBI. International VLBI experiments including the Kashima 26m antenna led to confirmation of the plate tectonics by recent geodetic measurements in 1980s (Heki et al. 1987; Matsuzaka et al. 1991). After transferred to GSI in 1992, the Kashima 26m antenna has been active in various international VLBI programs as the core site of Japan, contributing to the maintenance of International Terrestrial Reference Frame (ITRF) (http://lareg.ensg.ign.fr/ITRF/) and determination of EOP.

However, due to its inevitable aging after 34-years of operation, the Kashima 26m antenna was out of service from October 2002, and demolished at the end of March 2003. The role of Kashima 26m antenna has been inherited to the Tsukuba 32m antenna, which was established in 1998. VLBI observations to tie Kashima and Tsukuba were performed more than 7 sessions from 1998 to 2001.

One of the largest achievements of the Kashima 26m antenna is that it was used as the de facto origin of the Japanese Geodetic Datum 2000 (JGD2000), which superseded the Tokyo Datum in April 2002 with a revision of the Survey Act in June 2001.

4.2 Contribution to Japanese Geodetic Datum 2000

Land surveying and navigation require a geodetic reference frame, a combination of ground-based control points and a set of site coordinates of those points. With the introduction of space geodetic techniques such as VLBI and GPS that can measure the Earth from space, the world common geodetic reference becomes a reality for the first time.

Since April 2002, Japan has introduced a new geodetic reference system JGD2000, which is a geocentric system compliant to the International Terrestrial Reference System (ITRS) maintained by the International Earth Rotation and Reference Systems Service (IERS).

VLBI played a key role in calculating Geodetic Coordinates 2000, which is the official set of site coordinates of national geodetic control points that realize JGD2000 on the ground. First, the coordinates of the Kashima 26m antenna at epoch 1997.0 were calculated from the ITRF94 coordinates and velocities at epoch 1993.0 published in the IERS Technical Note No. 20 (Murakami and Ogi 1999). Positions of two mobile VLBI stations at Shintotsukawa in northern Japan, and Kainan in western Japan, were determined with respect to Kashima using the data from domestic VLBI observations. Please not that ITRF94 was the latest solution available from IERS when GSI started the preparation of JGD2000 in 1996.

Next, the coordinates of GPS-based control points (GEONET stations) were calculated using the 6-day GPS data by fixing the ITRF94 coordinates of three VLBI stations. In prior to that, local tie surveys between VLBI stations and nearby GEONET stations were conducted by GPS with a precision of a few cm.

Referring to the coordinates of GPS-based control points, coordinates of the first, second, and third-order control points (about 39,000) were calculated using the latest data available either by GPS, Electro Distance Meters (EDM), or theodolites. Coordinates of the fourth-order control points (about 64,000) were derived by a transformation program using interpolated transformation parameters of neighboring upper control points.

Because Geodetic Coordinates 2000 are now widely used for public surveys in Japan, extreme care should be taken for its maintenance because of large and complex crustal deformations in Japan. To validate coordinates and velocity of GEONET, high precision local tie surveys between the current VLBI stations and nearby GEONET stations have been conducted using a cateye reflector and a conventional surveying technique (Matsuzaka et al. 2002).

4.3 UT1- intensive series

Earth Orientation Parameters (EOP), i.e. precession, nutation, rotational velocity (UT1-UTC), and pole coordinates, connect a Terrestrial Reference System on the ground and a Celestial Reference System in the sky. EOP can be provided completely only by VLBI because it is the only technique that has direct access to the Celestial Reference System (Drewes et al. 2002). Such information is necessary for precise baseline analysis of GPS, performing orbit controls of space crafts, and decision to insert or delete a leap second in UTC. Furthermore, temporal variation of EOP sheds light on the dynamics of the inner Earth that can not be directly observed.

For a rapid turnaround of EOP, IVS conducts UT1-Intensive series (INT1) at the baseline Wettzell - Kokee Park. From July 2002, additional UT1-Intensive series named INT2 were carried out on the baseline Tsukuba - Wettzell (IVS 2002). Fig. 2 shows related baselines. An independent monitoring of UT1 becomes possible with INT2

organized by Bundesamt für Kartographie und Geodäsie, the Geodetic Institute of the University of Bonn, and GSI. GSI is responsible for a weekly 1-hour observation at Tsukuba and data processing with K4 correlator, submitting UT1 solutions to IVS each week (Miyagawa and Kurihara 2002).

Fig.3 compares recent results of INT1 and INT2 with the IERS C04 series subtracted as a reference. The IERS C04 is a combined solution of UT1 with various space geodetic techniques. INT1 and INT2 generally show good agreements with few exceptions. A standard deviation of each INT2 session is about 5 to 10 micro seconds, which satisfies IVS's requirement. However, there seems to be a few micro seconds bias with the IERS C04 series, which might suggest systematic differences between VLBI and other techniques.

From April 2003, INT2 series are conducted every Saturday. If they are fruitful, we will increase the number of INT2 series to two sessions per week in future. This is expected to increase accuracy and timeliness of UT1 products of IVS.

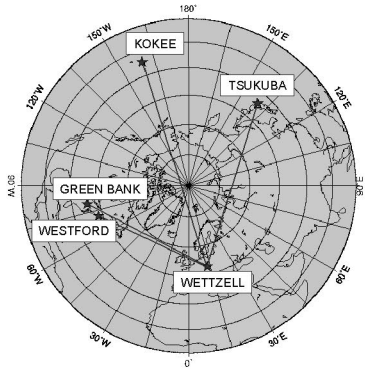


Fig. 2 VLBI baselines related to IVS UT1-Intensive series

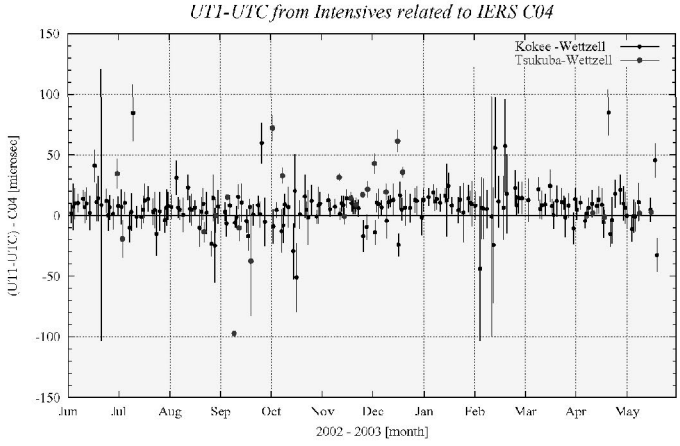


Fig. 3 UT1-UTC determined from IVS UT1- Intensive series. Larger circles denote the results from the Tsukuba-Wettzell baseline.

4.4 Advanced observation techniques

From 2002, GSI performed remote VLBI observations at domestic VLBI stations. Station equipment is remotely controlled by an expert stationed at Tsukuba except for the initial loading of recording tapes that are handled by local assistants. The remote observations contribute to the reduction of operational costs and increase of observation times. In 2003, monthly domestic JADE sessions are planned using the remote observation method.

A very high speed 2-Gbps optical cable has been connected to Tsukuba station under the joint research project with National Astronomical Observatory (NAOJ). An online VLBI experiment between Tsukuba 32m antenna and Usuda 64m antenna of NAOJ was performed in December 2002 (Takashima, Kurihara 2002). With an expansion of such a high speed optical network over Japan, a real-time online VLBI observation might be possible in future.

4.5 Global solutions

GSI has started a production of VLBI global solutions using the archived data at the IVS database. Fig. 4 shows velocities of VLBI stations nearby Japan using IVS data from 1980 to 2003. Such information is important for both the maintenance of JGD2000 and validation of GEONET time series.

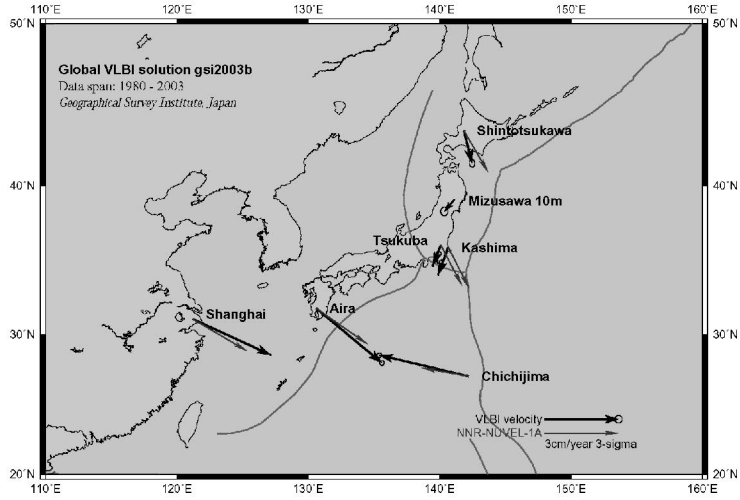


Fig. 4 Velocities of VLBI stations around Japan obtained from GSI's global analysis using IVS data from 1980 to 2003.

5 Conclusions

We have seen GSI's VLBI programs focusing on the synergy with GPS in Japan. Long term geodetic observations at Kahima were essential to the realization of ITRS in Japan as JGD2000. The inter-technique solution of VLBI and GPS increased reliability of this national geodetic infrastructure, which allows more efficient GPS surveying than the previous Tokyo Datum. A new challenge awaits Tsukuba - Wettzell UT1-intensive series towards a rapid turnaround of UT1, which will be beneficial for high precision GPS users.

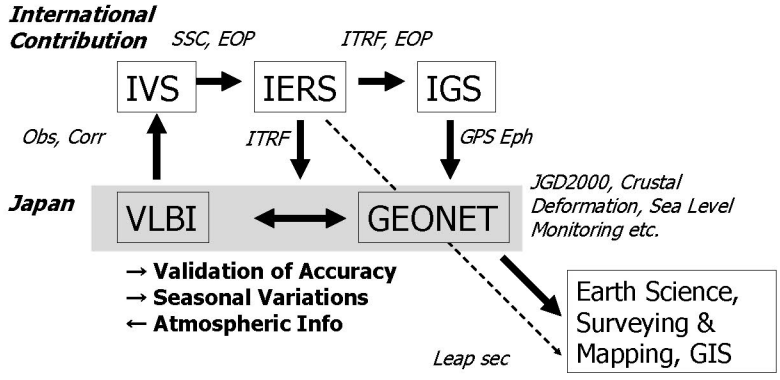


Fig. 5 Synergy of VLBI and GPS in Japan. SSC and Eph denote a Set of Site Coordinates and Satellite Ephemeris, respectively.

Fig. 5 summarizes the synergetic relationship between VLBI and GPS seen from Japanese point of view. As one of IVS core sites, the Tsukuba VLBI station contributes to a Set of Site Coordinates (SSC) and EOP solutions of IVS. The correlator at Tsukuba also contributes to IVS's UT1 solutions of the K-4 intensive series. IERS combines SSC and EOP solutions from each space geodetic technique. ITRF and EOP solutions from IERS are important constraints for precise orbit determination of GPS satellites at IGS. Back in Japan, IGS products are of great importance in analyzing daily GPS data from GEONET to monitor crustal deformations in Japan. GEONET is also a key infrastructure for surveying, mapping, and GIS application. To validate GEONET's accuracy, inter-technique comparison with domestic VLBI solutions is of great importance. The synergy between VLBI and GPS can open new frontiers of Geodesy such as an interpretation of ppb-level seasonal variations seen in GEONET time series. In turn, GEONET can provide atmospheric and ionospheric information over Japan which are helpful in domestic VLBI analyses. Because of these synergetic relationships, GSI continues its commitment to VLBI with collaboration with IVS, CRL, NAOJ, and other related organizations.

References

Drewes H, G Beutler, R Rummel (2002) Challenges for VLBI Within an Integrated Global Geodetic

Observing System, International VLBI Service for Geodesy and Astrometry 2002 General Meeting Proceedings, 24-32
Hatanaka Y (2002) Study on Optimal Analysis Strategy of Dense GPS Networks, Ph.D dissertation, submitted to Graduate School of Science, Hokkaido University
Heki K, Y Takahashi, T Kondo, N Kawaguchi, F Takahashi, N Kawano (1987) The relative movement of the North American and Pacific plates in 1984-1985, detected by the Pacific VLBI network, Tectonophysics, 144, 151-158
IVS (2002) IVS Newsletter Issue 4, December 2002
Matsuzaka S, M Tobita, Y Nakahori, J Amagai, Y Sugimoto (1991) Detection of Philippine Sea Plate Motion by Very Long Baseline Interferometry, GRL, 19, 8, 1417-1419
Matsuzaka S, Y Hatanaka, K Nemoto, Y Fukuzaki, K Kobayashi, K Abe, T Akiyama (2002) VLBI-GPS Collocation Method at Geographical Survey Institute, International VLBI Service for Geodesy and Astrometry 2002 General Meeting Proceedings, 96-100
Miyagawa K, S Kurihara (2002) Tsukuba VLBI Correlator, International VLBI Service for Geodesy and Astrometry 2002 Annual Report, 197-200
Murakami M, S Ogi (1999) Realization of the Japanese Geodetic Datum 2000 (JGD2000), Bulletin of Geographical Survey Institute, Vol. 45, 1-10
Takashima K, S Kurihara (2002) Tsukuba 32-m VLBI Station, International VLBI Service for Geodesy and Astrometry 2002 Annual Report, 157-160

Ground-based verification of four-way Doppler measurement for the SELENE mission

H. Noda[1], K, Asari[1] , N. Namiki[2], T. Iwata[3], Y. Kono[1] , H. Hanada[1] , N. Kawano[1], and Z. Yamamoto[4]

[1] Department of Earth Rotation, National Astronomical Observatory of Japan, Hoshigaoka 2-12, Mizusawa, Iwate, Japan
e-mail: noda@miz.nao.ac.jp
[2] Kyushu University, Fukuoka, Japan
[3] National Space Development Agency, Tsukuba, Ibaragi, Japan
[4] Institute of Space and Astronautical Science, Sagamihara, Kanagawa, Japan

Abstract. In Japan's lunar mission called SELENE (SELenological and ENgineering Explorer), detailed measurements of the lunar gravity field will be made by using two methods. The first is the differential VLBI measurement with two sub-satellites, and the second is the four-way Doppler measurement by using a sub-satellite which relays radio waves from the main lunar orbiter. The lunar far-side gravity field will be directly measured for the first time by using the second method, which will give us new information about the inner structure of the Moon. For the four-way Doppler measurement, the confirmation of the link establishment is an important issue because there is no telemetry relay from the main satellite to the sub-satellite in SELENE. The instruments are now under development, and a compatibility test of the instruments to the ground antenna station was carried out in March 2003. We have applied recently proposed algorithms for the confirmation of the link establishment and confirmed that they work successfully for data obtained during this test. The accuracy of the 2- and 4-way Doppler measurement are also measured and confirmed to be the expected level.

1 Introduction

There have been many studies about the determination of the near-side lunar gravity field since the first result by Lorell and Sjogren (1968). After the Clementine mission (Lemoine *et al.* (1997)), recent great progress was realized by the Lunar Prospector. Lunar gravity models of 100th-degree expansion of the spherical harmonics coefficient have been made (Konopliv *et al.* (2001)). Even with these extensive programs, however, the lunar far-side gravity field has been unresolved so far, because these data come from 2-way range rate tracking from the ground station so no direct measurement is possible when spacecraft is over the far-side. Besides, when we utilize the spherical harmonic analysis, the far-side gravity potential is only a result of mathematical calculation and it does not necessarily reflect the true potential (e.g. Floberghagen *et al.* (1999)). So, actual far-side observation is inevitably needed for the basic research of the physics of the Moon.

Japan's lunar explorer called SELENE (SELenological and ENgineering Explorer) (Sasaki *et al.* (1998)) is scheduled to be launched in 2005. Determination of the gravity field of the far-side of the Moon is one of the major science targets. The SELENE mission consists of one main orbiter and two sub-satellites called Rstar and Vstar (R stands for Relay, and V for VLBI). The SELENE trio has the same inclination of almost 90 degree to cover all the latitudes of the Moon. They will be launched as one satellite and the two sub-satellites will be released from the main orbiter. Finally, the main orbiter will be put into 100×100 km orbit, and Vstar into 100×800 km orbit, Rstar into 100×2400 km orbit, respectively. Thus, even when the main orbiter goes behind of the Moon, the relay sub-satellite will stay visible from the Earth, enabling it to relay the radio waves from the main orbiter to measure the far-side gravity field of the Moon by the Doppler tracking (RSAT mission). The gravity field around the lunar limb areas will also be measured by using Very Long Baseline Interferometer (VLBI) technique by using two radio sources onboard both Rstar and Vstar (VRAD mission). Combined with these data, the lunar "global" gravity field will be measured for the first time, and they contribute to the study of the inner structure of the Moon.

In the Doppler measurement, the Doppler shift from the radio wave whose frequency is precisely known is measured. Data from 1-way Doppler measurement will not be used as science data in this mission because it is less accurate for the science target. So, 2-way Doppler measurement is used for the precise measurement, where the downlink signals are coherent to the uplink radio wave which is made based on a very stable frequency standard with, for example, Hydrogen masers. In the SELENE mission, there is no telemetry (or command) relay from the main orbiter to Rstar and the link status between them is not known from the telemetry. Therefore, the real-time monitoring of the lock state monitoring is needed to achieve the mission. On the other hand, the uplink should be swept to re-establish the coherent link as soon as possible in case the link becomes unlocked, because the data taken during the unlocked period are not used as the

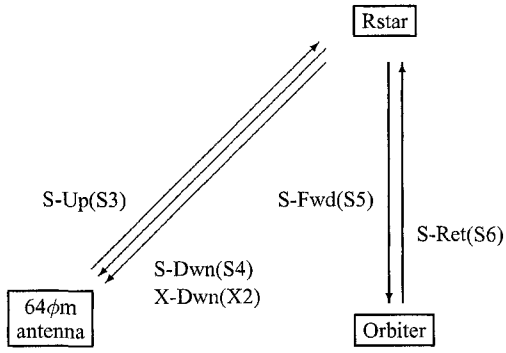


Fig. 1. The system of 4-way Doppler measurement.

science data. Furthermore, the assigned time for the switching of the ground antenna pointing from the main orbiter to Rstar is about 5 minutes, while the orbiter stays above the far-side about 1 hour. Therefore the time of decision whether the links are established or not should be short to avoid loss of data. In light of this situation, Asari *et al.* (2003) proposed a new method to monitor the status of the communication link between satellites which have no telemetry relay. During the final stage of the development phase of the RSAT instruments, we tested their algorithms to data obtained in a compatibility test of the instruments with the ground antenna station (Usuda Deep Space Center (UDSC) of the Institute of Space and Astronautical Science (ISAS)) held in early 2003, and the results are reported in this paper. The accuracy of the Doppler frequency measurements during this test is also briefly mentioned.

2 Principles of the coherent state monitoring in SELENE/RSAT mission

In SELENE RSAT mission, Doppler data from three links will be used for the lunar gravity field measurement. The first is the 2-way link between ground and the main orbiter when it is above the near-side of the Moon (main orbiter 2-way), and the second is the 2-way link between ground and Rstar (Rstar 2-way), and the last is the 4-way link which will be mentioned following. These data are used for the orbit determination and the estimation of the lunar gravitational potentials by GEODYN II software. Multiple frequencies used for Rstar 2-way and 4-way range and range rate measurements are listed in Table 1. The system of the 4-way measurement is shown in Figure 1. At first, the S-band uplink (S3) from the ground antenna goes to Rstar. The uplink is transmitted coherently to the ground antenna through S-Dwn(S4) for Rstar 2-way measurement. At the same time, the S/X-band transponder aboard Rstar forwards the uplink to the main orbiter (S5). An S-band transponder on the main orbiter returns this radio wave to Rstar (S6). An X-band frequency (X2) is made on Rstar by $X2$=680/221×S3+S6 and it is received at the ground antenna. When no signal is forwarded from Rstar to the main

link(name,power[dBm])	Freq.[MHz]	Coh. ratio
S-Up to Rstar (S3, -95)	2081.466	—
S-Down from Rstar (S,-139)	2260.415	241/220×S3
S-Fwd to Orbiter (S5, -134)	2241.579	238/221×S3
S-Ret from Orbiter (S6, -122)	2051.614	270/295×S5
X-Down from Rstar (X2,-132)	8456.125	680/221×S3+S6

Table 1. A list of frequencies and their relationship used in SELENE/RSAT mission.

orbiter, S6 is generated by an oscillator onboard the main orbiter. Once the S-band transponder onboard the main orbiter captures the forwarded signal, it generates return signal (S6) which is coherent to S3. Thus the X-band downlink (X2) becomes dependent on the S-band uplink (S3) in the coherent mode. For Rstar 2-way measurement, S-Dwn(S4) is modulated so that the telemetry of Rstar, including the lock state, can be monitored. On the other hand, no telemetry signals are added to S-Fwd (S5), S-Ret(S6), and X-Dwn(X2).

In Table 1, the nominal level of the power from the link designed in each link is also shown. Planer antennas aboard Rstar will be used in the forward link S5, and this is considered to be the only link where the 4-way link could be broken because it has the smallest margin.

In Asari *et al.* (2003), they concluded that the frequency stability was the best measure for the lock state monitoring because all the links are coherent to a Hydrogen maser on the ground antenna station in coherent mode, while the freerun state is made by a less stable crystal oscillator onboard spacecraft. The easiest way to judge the link status is to get second differences of the downlink frequency and see their amplitude. In addition, monitoring the response to the downlink by sweeping the uplink frequency with relatively slow sweep velocity ($\Delta f/\Delta t$ = 10 Hz/sec) and small amplitude ($\Delta f = \pm 20$ Hz) is also a useful method in case the carrier signal becomes weak. In SELENE, the response of the X-band downlink by the change of the uplink frequency results in $680/221\Delta S3+0\sim3\Delta f$ when the link is not coherent to the uplink, otherwise $680/221\Delta S3+\Delta S6\sim4\Delta f$ because S6 is also coherent to the S3 with ratio of 270/295. One can judge the coherency according to the difference of the amplitude.

3 Test settings and data analysis

3.1 Test configuration

For the compatibility test, the flight models were reconfigured to a simulator of Rstar (which consists of an S-band transponder and an S/X-band transponder) and of the main orbiter (an S-band transponder). These simulators were connected with cables and attenuated uplink signals were fed by a cable to the Rstar simulator. Downlinks of both S- and X-band from the Rstar simulator were also put to the ground receiver system via cables. Counts by the Doppler counter of the ground system were obtained in every 0.1 second, which were time-integrated to calculate Doppler frequencies with any wished time period as the post-processing. The 64m ϕ antenna at UDSC was pointed to the sky to reduce the system noise temperature ($T_{SYS} \sim 70$ [K]), except for short times during the 4-way total test when the antenna was pointed to

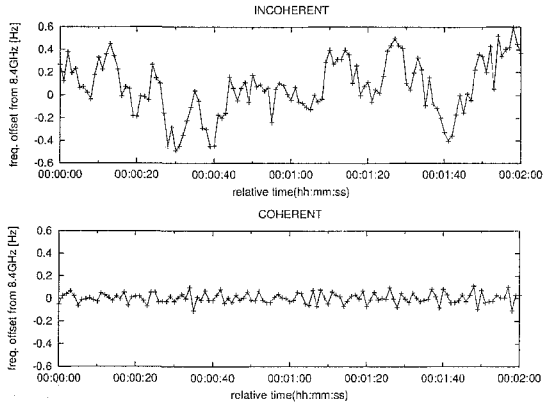


Fig. 2. Time series of the X-band downlink (0.1-second values are integrated to 1-second value). Upper and lower panel are data of incoherent and coherent mode, respectively. Horizontal scale is 2 minutes, and vertical axis show the difference from the averaged value. Correct evaluation is hard to be done within a few seconds though the difference becomes clear within two-minute scale. The periods when these data are taken are shown by arrows in Figure 3.

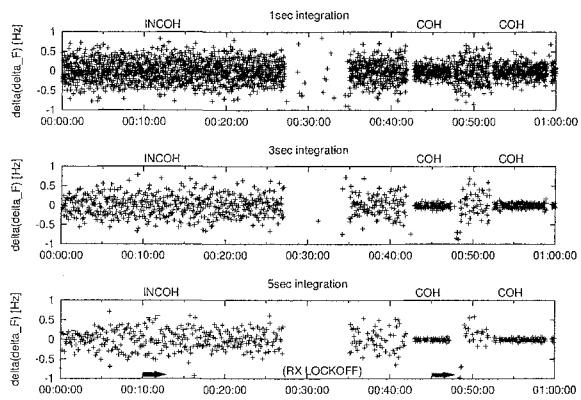


Fig. 3. Integration time dependency of second differences of the Doppler frequency, 1, 3, and 5 second integration from the top panel. Full horizontal axis corresponds to an hour. The 4-way link is established between 00:40:00-00:43:00 and 00:50:00-01:00:00 (labeled as "COH") Five-second integration is enough to identify coherent/incoherent, although it takes more time than we had expected. Between 00:27:00-00:35:00 the receiver of the ground system is locked off (labeled as "RX LOCKOFF"). Arrows in the bottom figures correspond time intervals appeared in Figure 2. Data was taken on 4th March in the connection test.

the Moon to check the increase of the system noise ($T_{SYS} \sim$ 300 [K]).

The tests were carried out as follows. At first, after the cable setting including level check, connection confirmation tests were done. In these opportunities, the first 4-way Doppler data with relatively strong power were obtained. After testing the threshold levels of capturing and sustaining of the link for the each link, end-to-end tests of 2-way range and range rate and 4-way range rate measurement were held. The second data set were obtained at this time. Tests of slow frequency sweep of the uplink were done in both coherent and incoherent state at this time.

3.2 Data presentation

Figure 2 shows a comparison between incoherent (upper panel) and coherent (lower panel) mode of the X-band downlink, respectively. These data were taken during the connection test on March 4th, 2003 when the uplink and relay power were high enough so that the energy could be conveyed between links with enough margins. The horizontal scale is 2 minutes, and 0.1-second values are integrated to 1-second value in each data point. The vertical axes show the difference from the averaged frequency. From this figure, one can say that the difference of the frequency stability is apparent between these two cases for two-minute scale. In short time (*i. e.* $\sim$ seconds), however, the difference is not necessarily clear.

Figure 3 shows a comparison of rates of the Doppler rate, namely, second differences of the X-band downlink frequency for 1-, 3-, and 5-second integration data, respectively. This measure, which is described as a useful measure for lock state monitoring in Asari *et al.* (2003), is defined as $(f_{i+2} - f_{i+1}) - (f_{i+1} - f_i) = f_{i-2} - 2f_{i+1} + f_i$, where f_i denotes the time-integrated frequency from t_{i-1} to t_i. Time intervals appeared in Figure 2 correspond to the arrows in the bottom panel. From the difference of the amplitude appeared in the bottom panel (5-second value data), it is known that the data of 00:44:00-00:47:00 and 00:50:00-01:00:00 are in the coherent mode because the amplitude of the second differences of the frequency are much smaller than those of the ambient. To evaluate the frequency stability, we calculated the Allan standard deviation (Figure 4). In this figure, the horizontal axis is accumulation time in log scale, and the vertical axis is Allan standard deviation. A line labeled as "COH" is data in coherent, and "INCOH" are of incoherent data. At one second at the horizontal axis, for example, the difference between incoherent and coherent is less than one order of magnitude while data differ one order for 5-second data, and the evaluation of whether all the links are coherent or not is not easy. It is also clear in Figure 3.

Figure 5 is a result of the slow sweep test. The horizontal axis corresponds to 40 second, and each data are 0.1 second raw data. The vertical axis is the frequency offset from the standard level when no sweep is done. In this figure, lines with large amplitude is data of coherent mode, and small one comes from the incoherent mode. The coherency can be ap-

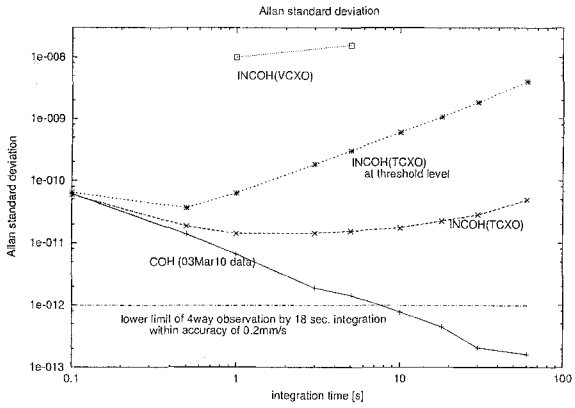


Fig. 4. Comparison of Allan standard deviation of incoherent (labeled as "INCOH(TCXO)") and coherent("COH") state. One of the incoherent mode data shows values worse than another because it was obtained in a situation where the forward link from Rstar to the main orbiter was just below the threshold level. 1.0×10^{-12} is the required stability of the downlink for 4-way observation with an accuracy of 0.2 mm/s Doppler measurement with 18 second integration, which is enough value to achieve the science goal. An incoherent data based on a VCXO is also plotted on the same figure (see text).

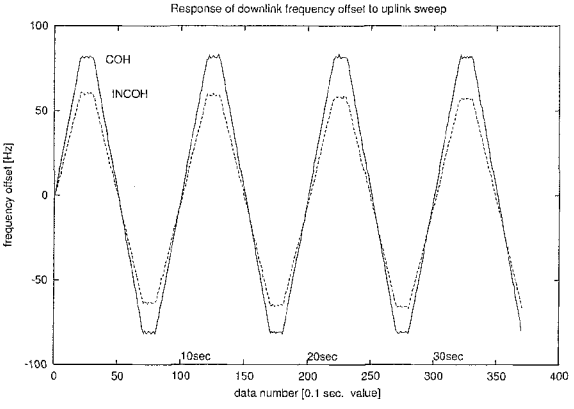


Fig. 5. Response of the X-band downlink to the slow sweep of the S-band uplink (Δf/Δt=10Hz/s, Δf=$\pm$20Hz) for both coherent and incoherent cases. The vertical axis is the frequency offset and the horizontal axis is data number of 0.1-second value. The incoherent and coherent mode can be identified with the difference of the amplitude, namely, the larger amplitude corresponds to the coherent mode.

parently judged by the difference of amplitude with 20 Hz (Δf).

3.3 Accuracy of Doppler measurements

According to the previous estimation, the accuracy of range rate measurements for 2-way (S-band) and 4-way (X-band) are 0.1 mm/s and 0.36 mm/s, respectively, when we only consider the thermal and phase noise. 23 measurements were made for both 2-/4-way Doppler frequency by changing the C/N_0 of each links during the test, and the deviations of the range rate data for 2-/4-way ranged within 0.05 $\sim$ 0.1 mm/s and 0.17 $\sim$ 0.32 mm/s, respectively. The required accuracy of the ranging of Rstar is 20 m, while the result was 0.5 $\sim$ 4.8 m, and the data were confirmed to satisfy the required level.

4 Discussion and concluding remarks

Concerning the integration time for the coherent status monitoring, we conclude that five-second integration is needed to distinguish coherent from the incoherent mode, because the difference of Allan standard deviation becomes more than one order in Figure 4, combined with visual inspection of Figure 3. As a result, more integration time was needed than the result by Asari *et al.* (2003), where they concluded that 1-second integration would be enough to distinguish between coherent and incoherent with more than 2 order difference in Allan standard deviation. This may come from the ill-estimation of the phase noise of the crystal oscillator. In terms of separating coherent mode from the incoherent mode, however, the link monitoring algorithms work well by selecting appropriate integration time.

In the actual case, however, other noise sources such as (i) increase of the system noise temperature to 300K by pointing the ground antenna to the Moon, (ii) change of the S/N ratio due to the change in distance between the satellites (iii) media effects by the ionosphere and the troposphere, and (iv) frequency dynamics due to the orbit motion of satellites relative to the Earth and between them, will be added. The increase of the system noise temperature (i) was confirmed as 5.7 dB when we pointed the antenna to the Moon during the test. This means that the threshold level of the satellite acquisition will be increased by 5.7 dB. From tests in which the link level was varied, we have confirmed that the Allan standard deviation is not severely affected by the change of S/N ratio unless the level becomes near the threshold level. In terms of S/N ratio, (i) and (ii) are the same things, and they do not influence the lock state monitoring unless the link level goes down to the threshold level. The ionospheric effect and the frequency dynamics both could affect the lock state monitoring algorithm because they affect the frequency stability. In order to obtain the actual data, Asari *et al.* (2003) calculated the Allan standard deviation of NOZOMI spacecraft which is a spin-stabilized satellite with the spin period of 8 second in a planetary orbit between the Earth and Mars. According to them, the long-period term of the Allan standard deviation of NOZOMI becomes much larger than the estimated values, and they conclude that this is because the estimation and subtraction of the orbit motion is not incomplete. For the

monitoring, we will not use the long-period term, so this may not be a problem. In the short-period, for example, the Allan standard deviation at 10 second is 1×10^{-12} and it is still somewhat larger by factor compared to the test. This may come from mainly the ionospheric effect, but the difference from the incoherent mode is still larger by two order and the monitoring will be done correctly.

The orbit motion of the satellites would make Doppler shift of 150 kHz from the base frequency. This value is not included in the compatibility test, and in the actual case, the evaluation of the link status becomes more difficult for the slow sweep method than the compatibility test. In reality, this Doppler shift should be subtracted by using predicted orbit data or by extrapolating Doppler data obtained before the observation time. In this case, the accuracy of the determination of the Doppler frequency originated from the orbit motion must be smaller than the difference of the downlink response of incoherent and coherent mode, namely, Δf (20 Hz).

One of the goals of the mission is to measure the lunar gravity field with one order precision in 2-30th coefficients of spherical harmonic expansion compared to the Lunar Prospector models (Matsumoto *et al.* (2003)). In the simulation study, the assumed level of the accuracy of the Doppler measurement is 0.2 mm/s with 18-second integration for 4-way Doppler measurement. Although the compatibility test corresponds to the case when only thermal and phase noises of the links are taken into account, the result of the test satisfies this standard within 8-second integration as shown by a dotted line in Figure 4. The error sources listed above will be added to the test result in the real operation. Major causes of the Doppler error are the thermal and phase noise of the front end and the delay by the ionosphere. The estimated values of range rate error are 0.36 mm/s and 0.33 mm/s for the thermal/phase noise and the ionospheric effect for the X-band, respectively, and the root sum square becomes 0.5 mm/s, and this value will be a reasonable value for the actual case.

The final remark is about the onboard instrument itself. As of March 2003, the instrument design is almost end before the compatibility test. However some parts still need to be fixed. A major change of the hardware was to remove a temperature compensated crystal oscillator (TCXO) on the main orbiter which was to be used as the source oscillator of freerun state, because of an electromagnetic interference between the phase-locked loop (PLL) circuits of the S-band transponder. Instead, the same voltage compensated crystal oscillator (VCXO) in the PLL circuit is used as the freerun source. In terms of the lock state monitoring, however, it turned out to be a merit because the frequency stability was almost two orders worse than the TCXO and the state could be judged in shorter times than of the case when a TCXO was used (see the Allan standard deviation of "INCOH(VCXO)" in Figure 4). The algorithms themselves will not be affected and only the integration time will be changed to a shorter time (1second). The selection of the integration time will be fixed during the commissioning phase of two months after the lunar orbit insertion of SELENE.

Acknowledgements. The authors thank W. M. Folkner and an anonymous referee for evaluating this paper. The instruments in RSAT/VRAD mission are developed by the NEC-TOSHIBA Space System Co. Ltd. . A reference data of VCXO stability in Figure 4 was obtained and offered by them.

References

Asari, K., et al., (2003) New methods to monitor a Satellite-to-Satellite 4-way link (in Japanese), *IEICE Trans.*, **J86-B**, No. 11, 2328–2335

Floberghagen, R. K. et al. (1999) Regularisation of lunar gravity field solutions, *Adv. Space Res.*, **23**(11), 1801–1807

Konopliv, A. S. et al. (2001) Recent gravity models as a result of the Lunar Prospector mission, *Icarus*, **150**, 1-18

Lemoine, F. G. R. et al. (1997) A 70th degree lunar gravity model (GLGM-2) from Clementine and other tracking data, *J. Geophys. Res.*, **102**, 16339–16359

Lorell, J. and W. L. Sjogren (1968) Lunar gravity : preliminary results from Lunar Orbiter, *Science*, **159**, 625

Matsumoto, K. et al. (1999) Impact of far-side satellite tracking on gravity estimation, *Adv. Space Res.*, **23**(11), 1809–1812

Matsumoto, K. et al. (2003) Anticipated accuracy of lunar gravity field model from RSAT/VRAD mission, *Abstract of IUGG 2003.*

Namiki, N. et al. (1999) Selenodetic experiments of SELENE: relay satellite, differential VLBI, and laser altimeter, *Adv. Space Res.*, **23**(11), 1817–1820

Sasaki, S. et al. (1998) SELENE Mission: Scientific Objectives (abstract) , *Proc. 21st Inter. Sympo. Space Tech. Sci.*, 98-i-01

Results of the Critical Design of RSAT/VRAD Mission Instruments on SELENE Sub-satellites Rstar/Vstar for Selenodesy

T. Iwata, T. Sasaki, T. Izumi
Japan Aerospace Exploration Agency, 2-1-1 Sengen, Tsukuba, Ibaraki 305-8505, Japan

Y. Kono, H. Hanada
National Astronomical Observatory, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

N. Kawano, F. Kikuchi
Earth Rotation Division, National Astronomical Observatory,
2-12 Hoshigaoka, Mizusawa, Iwate 023-0861, Japan

Abstract. Four-way Doppler measurements and differential VLBI will be executed by SELENE to obtain highly accurate global mapping and improve the model of lunar gravity. Two small sub-satellites: Relay Satellite (Rstar) and VLBI Radio Satellite (Vstar), which are separated from SELENE Main Orbiter, perform these selenodetic observation using Relay Satellite Transponder (RAST) and VLBI Radio Sources (VRAD). These sub-satellites have no thrusters to control orbits and attitude to yield precise measurements of orbits perturbed by gravity anomaly. Lack of thrusters, however, causes the instability of attitude that interfere the selenodesy observation. The tip off at the separation and the solar radiation pressure torque dominantly affect the attitude, hence, we evaluated them in the critical design phase of the sub-satellite development. Properties of the newly developed release mechanism were obtained by ground tests, and the satellite attitude inclination by disturbances was analyzed. These results show that the design of the sub-satellites has enough properties to produce selenodesy data.

Keywords. SELENE, Moon, selenodesy, gravity, RARR, VLBI

1 Introduction

Precise determination of orbits of lunar probes perturbed by mass distribution allows determination of the gravity field which can be observed using a spherical surface harmonic expansion. Orbits of the probes were hitherto observed by two-way ranging and ranging rate (RARR) measurements of the relayed signals between ground stations and spacecrafts. Accumulated data of RARR measurements from Lunar Orbiter to Lunar Prospector have yielded the best current lunar gravity field model LP165P (Konopliv et al. 2001). The lunar gravity map is, however, less accurate above the lunar far side, because the lunar rotation and moon-to-earth revolution are synchronized so that most of the orbits above the far side have been estimated from the accumulated perturbation observed above the near side. The accuracy above the lunar rim is also inferior because the orbital perturbations across the rim are scarcely detected by RARR methods. We, therefore, plan to improve the gravitational model by combining four-way Doppler measurements and VLBI observations for the lunar spacecraft.

SELENE (Selenological and Engineering Explorer) is a lunar probe which will be launched by Japan Aerospace Exploration Agency (JAXA) in 2005. It is composed of Main Orbiter, and two small sub-satellites; Relay Satellite (Rstar) and VLBI Radio Satellite (Vstar) (Iwata et al. 2002). Rstar and Vstar will be separated from Main Orbiter and injected into the initial polar orbits of 2,400-100 km and 800-100 km in altitude, respectively. Four-way transmission between Main Orbiter and the ground station via Rstar will supply the Doppler measurements of the orbits above the lunar far side (Namiki et al. 1999). Radio sources on both Rstar and Vstar will deliver carrier waves for multi-frequency differential VLBI observations (Hanada et al. 2002). We report results of the critical design for these selenodesy missions.

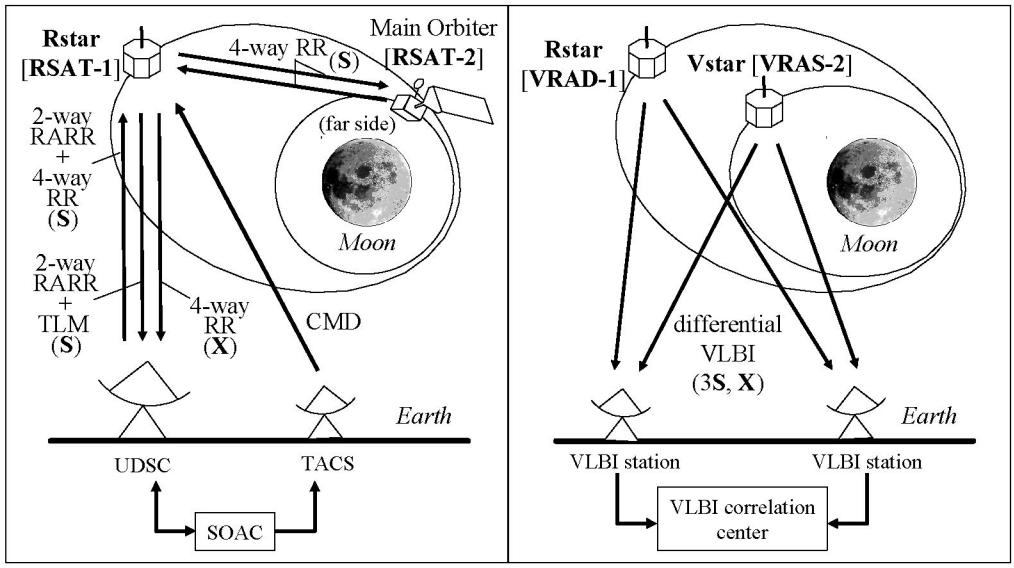


Fig. 1 Concepts of RSAT: Relay Satellite Transponder mission (left), and VRAD: Differential VLBI Radio Source mission of SELENE. Notes- RARR: ranging and ranging rate, RR: ranging rate, TLM: telemetry, CMD: command, S: S-band, X: X-band, UDSC: Usuda Deep Space Center, TACS: Tracking and Control Station, SOAC: SELENE Operation and Analysis Center.

2 Outline of the Mission

2.1 Four-way Doppler Measurements by RSAT

Relay Satellite Transponders (RSAT) on Rstar and Main Orbiter will perform the four-way Doppler measurements for Main Orbiter aviating above the lunar far side (Figure 1: left). The orbit of Rstar will be simultaneously measured by two-way RARR. Transponders on Rstar (entitled RSAT-1) relay both ranging signals for two-way RARR and carrier waves for four-way links. A transponder on Main Orbiter (RSAT-2) relays the four-way links.

S-band transponder of RSAT-1 receives ranging signals of S-band transmitted from 64-m antenna at Usuda Deep Space Center (UDSC), and returns them to UDSC to conduct two-way RARR. The transponder also extracts carrier waves from the received signals and relays them toward Main Orbiter. RSAT-2 on Main Orbiter receives the forward signals and returns them to Rstar. S/X-band transponder of RSAT-1 receives the returned signals, converts the frequency to X-band, and relays them to UDSC. The compositions of each component and the strategy for four-way link have been presented by Iwata et al. (2001).

Four-way Doppler measurements using RSAT will provide the first direct determination of a low lunar orbit above the far side. Analysis of the coverage considering the four-way links shows that our methods will present fully covered gravity map (Matsumoto et al. 2002). The accuracy of four-way Doppler and two-way RARR measurements have been examined by ground station compatibility tests and argued by Noda et al. (2003).

2.2 Differential VLBI Observation by VRAD

Differential VLBI Radio Sources (VRAD) on Rstar and Vstar will generate signals for multi-frequency differential VLBI observations (Figure 1: right), which provides the highly accurate positioning of both the satellites. Radio sources on Rstar (VRAD-1) and Vstar (VRAD-2) transmit three pairs of S-band and a pair of X-band carrier signals to VLBI radio telescopes where the narrow band-width VRAD terminal is installed.

The distribution of coherent S/X-band frequencies is chosen to calibrate the delay by the terrestrial ionosphere and to solve uncertainties of the delay over one wave-length. The accuracy of the satellite positioning by VRAD is estimated to be about 20cm, which is more precise by three orders than those obtained by hitherto two-way RARR methods (Hanada et al.

2002). Because Rstar/Vstar have no altitude and attitude maneuver by thrusters, orbital determinations with long arcs will present the highest sensitivity for the lower degrees of lunar gravity coefficients. They also improve the accuracy of a gravity map on the lunar rim (Matsumoto et al. 2002).

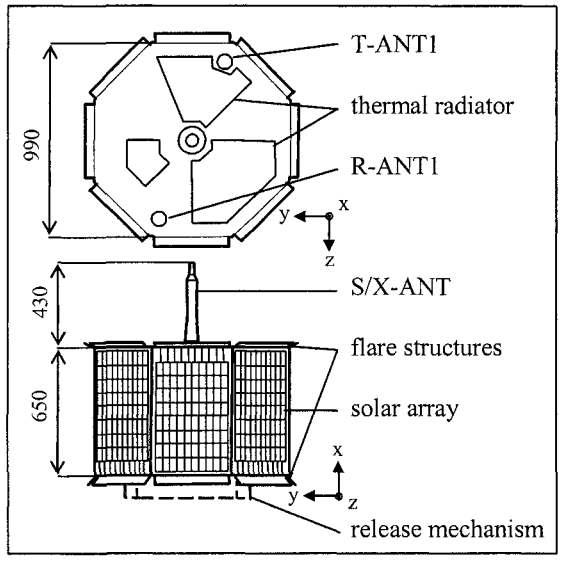


Fig. 2 Appearances of SELENE Relay Satellite (Rstar.).

Table 1. Mass properties of Rstar and Vstar.

parameter	value
total mass [kg]	47.9
centroid coordinates [m]	(0.325, 0.0 0.0)
MOI [kg m^2]	Ixx = 3.924
	Iyy = 3.749
	Izz = 3.743
	Ixy = 2.8060×10^{-4}
	Iyz = -1.1838×10^{-2}
	Izx = 2.4299×10^{-4}

Table 2. Surface characteristics of Rstar and Vstar.[a]

material	Cd	Cs	Ca
solar array	0.22	0.02	0.76
Ag Teflon	0.13	0.79	0.08
conductive MLI	0.15	0.00	0.85
Ge coating[b]	0.40	0.00	0.60

[a] Cd: diffusion reflectance, Cs: specular reflectance, Ca: absorptance. [b] On S/X-ANT, flare structures.

3 Design of the System

3.1 Overview of Mission Instruments and Bus Equipments on Rstar/Vstar

Rstar and Vstar are spin-stabilized small sub-satellites whose purpose is restricted to selenodetic observations. Configurations of Rstar are displayed in Figure 2, in which most parts of the release mechanism remains on Main Orbiter after the separation. Both Rstar and Vstar have octagonal prism main bodies which measure 1m×1m×0.65m. Mass distributions of Vstar are arranged to be the same with those of Rstar using mass dummy structures. Mass properties of these sub-satellites are shown in Table 2 with the coordinate system displayed in Figure 2. The total mass is 47.9 kg for each sub-satellite.

S/X-band coaxial vertical dipole antennas; S/X-ANT on Rstar/Vstar link signals toward ground stations with toroidal beams of 32 degrees width. Two pairs of S-band patch omni-directional antennas on Rstar link signals toward Main Orbiter with conical beams of 140 degrees width. Each pair of antenna for receiving (R-ANT1 and 2) and transmission (T-ANT1 and 2) on both upper and lower decks are settled in axial symmetries to reduce the Doppler shift on the data caused by the satellite spinning.

Thermal radiators of Ag Teflon on the upper deck discharge heat of the system, of which areas are individually optimized for each satellite. Most parts of main body surfaces except antennas and radiators are covered with conductive multi layer insulator (MLI). Table 2 displays diffusion reflectance: Cd, specular reflectance: Cs, and absorptance: Ca for each principal material on the surface.

3.2 Design of the Release Mechanism and Attitude Control

Neither of the sub-satellites has thrusters to control orbits and attitudes, which yields precise measurements of orbital perturbations and enlightens mass of the satellite system. Lack of thrusters, however, causes the instability of satellite attitude which interferes in the detection of gravity anomaly and reduce the mission period. The initial attitude conditions are settled by the

mechanism to release sub-satellites from Main Orbiter, then the nutation caused by the separation is removed by a nutation damper. We also adopted flare structures to maintain the attitude against solar radiation pressure. It is, hence, important to design the release mechanism, the nutation damper, and the flare structures to supply enough stable attitudes.

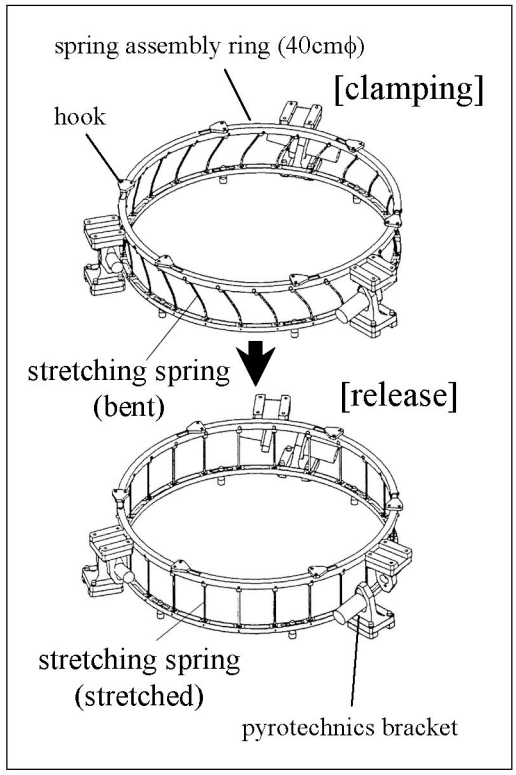


Fig. 3 Appearances of the release mechanism of SELENE sub-satellite during clamping (top) and after released (bottom).

We have developed a release mechanism with low mass of 4 kg. Figure 3 shows appearances of the release mechanism during clamping (top) and after released (bottom). Two spring assembly rings are connected with 24 stretching bow springs which bend during clamping and stretch after the release. Three sets of pyrotechnics separate the brackets on Rstar/Vstar and Main Orbiter. Hooks suspended on the cavities on the spring assembly ring supply the spin. An MLI cover assembly, which is abbreviate in Figure 3, is equipped on the exterior of the diameter of the spring assembly ring. The release properties of the mechanism are confirmed as described in section 3.3.

Nutation dampers are installed on the lower deck panel of Rstar/Vstar. They reduce nutation after the separation of the sub-satellites from Main Orbiter, which decrease the disturbance of the orbital perturbation detections. Flare structures with germanium coating attach to the edge of upper and lower decks (Figure 2). Sizes and surface characteristics of them are adjusted to balance solar radiation pressure with those on S/X-ANT (Table 2).

3.3 Ground Tests or Release and Attitude Properties

We confirmed the release properties by ground tests using breadboard model and engineering model of the release mechanism. A dummy mass structure of Rstar/Vstar was suspended by the ground equipment with 4m-length shock cord which is adjusted to balance the tension with the terrestrial gravity to imitate weightlessness (Figure 4). Basic release properties are; separation velocities along vertical line, horizontal velocities, spin rates, and tip off rates which are the rotation velocities perpendicular to spin the axis arising only in the separating process. Separation velocities and tip off rates were measured by the laser shiftmeters, and spin rates and horizontal velocities were analyzed from the image processor with video cameras. The release timings were observed by the gravity sensors. Because characteristics of the stretching springs are changed by a function of temperature and stress, release properties before and after expected thermal history on the ground and on orbit were obtained.

Results of the ground tests before and after thermal history are shown in Figure 5 which shows the distributions of spin rates and separation velocities. Because the release properties are sensitive to the moment of inertia (MOI) across the spin axis, we investigated the sensitivity to change in the mass distribution of the dummy mass structure. Properties with nominal and maximum MOI are denoted by circles and squares in Figure 5, respectively. Open and closed marks denote before and after thermal history, respectively. An area surrounded by a hexagon indicates the envelope of values with 3σ dispersion. Dashed lines indicate the limits of specific values as explained

in the next paragraph.

The lower limit of the separation velocity of 110 mm/s, and the upper limit of the spin rate of 108°/s are restricted to avoid collision between fabrics on each satellite with the values of 3σ dispersions. The lower limit of the spin rate of 60°/s is arranged with the values of tip off rate. We designed the ratio of the tip off rate to spin rate to be less than 0.1 which correspond to the nutation magnitude of 5.7°. The maximum obtained ratio was 0.06. These results show that parameters of release properties satisfy requirements for the secure separation and spin stabilization.

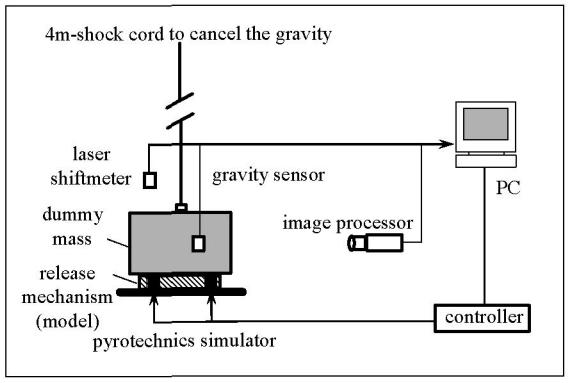


Fig. 4 A schematic configuration of the ground tests for release property measurements.

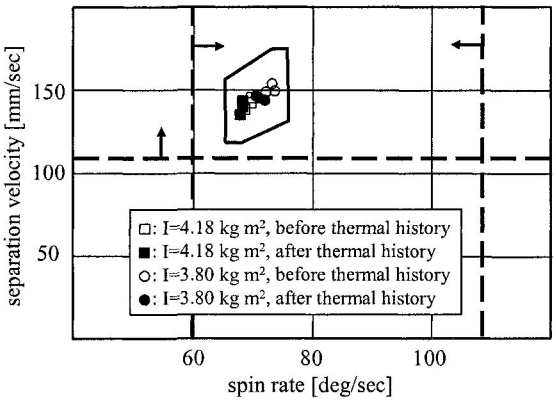


Fig. 5 Spin rate and separation velocity by the release mechanism obtained before and after thermal history in the ground tests. Marks and lines are denoted in the text.

4 Analysis for attitude properties of Rstar/Vstar

The tip off, which is a rotation perpendicular to the spin axis in the separation process, causes nutation on sub-satellites. After the nutation damper remove the tumbling of the attitude, the mean angle of the spin axis at the initial nutation remains as a tilt of the spin axis. The tilt cannot actively be diminished because of the absence of attitude control thrusters. The attitude is also inclined by torques of solar radiation pressure and gravitational gradient. We have estimated the properties of sub-satellites attitudes and evaluated them for selenodesy mission.

Table 3. Results of attitude analysis for Rstar.

factor	budget
slant of inertia axis	0.5°
tip off at separation	5.7°
attitude error at separation	1.0°
total satellite inclination [a]	13.6°
parallax of satellite orbit	0.7°
parallax of ground station	1.0°
total attitude error	15.3°

[a] Including the tilt by solar pressure and gravitational gradient torques.

Requirements for the attitudes of Rstar/Vstar are regulated by the beam pattern of S/X-ANT to link signals toward ground stations, and by the incident angle of the solar light to generate demand electric power. We define the total attitude error as the angle between the spin axis and the vertical line of the lunar orbital plane, which is the sum of the total satellite inclination, parallax of satellite orbit, and parallax of ground station. The total error is required as less than 16° during the mission period of one year.

The initial attitude is given by synthesis of the slant between inertia axis and mechanical axis, the tilt caused by the tip off at the separation, and the attitude error at the separation. The slant of inertia axis is determined to be less than 0.5° by results of the ground test for mass properties measurements using a mechanical test model of the sub-satellites. The tilt angle caused by the tip of at the separation has been estimated to be less than 5.7° by the ground tests of the release mechanism. The attitude error caused by the Attitude and Orbit Control Subsystem of Main Orbiter is designed as less than 1°.

Then the attitude analysis has been modified from the results in Ping et al (2001) using

parameters revised in the critical phase; mass properties, surface characteristics, structures, and release properties. The torque of solar pressure is a dominant factor for the inclination of Rstar/Vstar. It is derived from the surface properties (Table 1) assuming 10 % larger by uncertainties. The torque of gravitational gradient is obtained from the mass properties in Table 2. The sum of the satellite inclination is numerically analyzed and the result for Rstar is summarized in Table 3. The total attitude error is calculated to be 15.3°, which conforms to the design requirements for communication links and power supply.

The spin and inclination of the satellite raise the motion of antennas which causes the Doppler shift on RARR data and interferes in the detection of gravitational perturbations. Kikuchi et al. (2003) simulated RARR data interfered by the spin and inclination variation of a spin satellite using the results of Rstar/Vstar design. They have demonstrated that the spin and tumbling motion of a spin satellite can be estimated and calibrated on RARR data by least squares methods. This indicates that the methods are useful for the attitude estimation in the Rstar/Vstar operation and for the data calibration of RSAT/VRAD mission because the sub-satellites have no telemetry for their attitude.

Summary

The mission instruments and bus equipments for four-way Doppler measurements and differential VLBI by SELENE have been developed in the critical design phase. Two small sub-satellites; Rstar and Vstar have no thrusters to control orbits and attitudes to yield precise measurements of orbits. Nutation dampers and flare structures are adopted to reduce the inclination of the spin axis. The attitude instability of Rstar/Vstar is dominantly caused by the tip off at the separation and the solar radiation torque. We estimate the tilt angle after the separation from the properties obtained by the ground tests of the release mechanism. We analyzed the attitude inclination by disturbances using mass properties, surface characteristics, structures, and release properties. We, then, confirmed that the performances of the attitudes are satisfied with the requirements from the secure separation and spin stabilization to perform selenodesy mission.

Authors are grateful to Dr. N. Namiki of Kyushu University, Prof. T. Takano of JAXA; and members of RSAT/VRAD mission group.

References

Hanada H, Iwata T, Kono Y, Matsumoto K, Tsuruta S, Ishikawa T, Asari K, Ping J, Heki K, Kawano N (2002) VRAD Mission: Precise Observation of Orbits of Sub-satellites in SELENE with International VLBI Network, In: Vandenberg NR, Baver KD (eds) Proc. International VLBI Service for Geodesy and Astrometry 2002 General Meeting, NASA/CP-2002-210002, pp 72-76

Iwata T, Takahashi M, Namiki N, Hanada H, Kawano N, Heki K, Matsumoto K, and Takano T (2001) Mission Instruments for Lunar Gravity Measurements using SELENE Sub-satellites, J Geod Soc Japan 47: 558-563

Iwata T, Sasaki T, Izumi T, Kono Y, Hanada H, Kawano N, Takano T (2002) Properties of the Release Mechanism of SELENE Relay Satellite and their Influences for the Lunar Gravity Observation, In: Ishii N et al. (eds) Proc. 23rd International Symposium on Space Technology and Science. Sanbi Insatsu, Tokyo, pp 1928-1933

Kikuchi F, Kono Y, Hanada H, Iwata T, Ono T, Kawano N (2003) Attitude Estimation for a Spin Stabilized Spacecraft from Doppler Shift (in Japanese), Transact IEICE (B) J86-B, 6: 959-968

Konopliv AS, Asmar SW, Carranza E, Sjogren WL, Yuan DN (2001) Recent Gravity Models as a Result of the Lunar Prospector, Icarus 150: 1-18

Matsumoto K, Heki K, Hanada H (2002) Global lunar gravity field recovery from SELENE, In: Vandenberg NR, Baver KD (eds) Proc. International VLBI Service for Geodesy and Astrometry 2002 General Meeting, NASA/CP-2002-210002, pp 381-385

Namiki N, Hanada H, Tsubokawa T, Kawano N, Ooe M, Heki K, Iwata T, Ogawa M, Takano T (1999) Selenodetic Experiments of SELENE: Relay Satellite, Differential VLBI, and Laser Altimeter, Adv Space Res 23: 1817-1820

Noda H, Asari K, Iwata T, Kawano N, Yamamoto Z (2003) Ground-base Verification of Four-way Doppler Measurement for the SELENE Mission, this issue

Ping J, Sengoku A, Nagaoka N, Iwata T, Matsumoto K, Kawano N (2001) How Solar Radiation Pressure Acts on RSAT and VSAT with a Small Evolving Tip-off in SELENE, Earth Planets Space 53: 919-92

Application of a PZT telescope to In situ Lunar Orientation Measurement (ILOM)

H. Hanada, K. Heki, H. Araki, K. Matsumoto, H.Noda, N. Kawano
Division of Earth Rotation,
National Astronomical Observatory, 2-12 Hoshigaoka, MIzusawa-shi, Iwate 023-0861, Japan

T. Tsubokawa, S. Tsuruta, S. Tazawa, K. Asari
Mizusawa Astrogeodynamics Observatory,
National Astronomical Observatory, 2-12 Hoshigaoka, Mizusawa-shi, Iwate 023-0861, Japan

Y. Kono
Division of Radio Astronomy,
National Astronomical Observatory, 2-21-1 Osawa, Mitaka-shi, Tokyo 181-8588, Japan

T. Yano, N. Gouda
Division of Astrometry and Celestial Mechanics,
National Astronomical Observatory, 2-21-1 Osawa, Mitaka-shi, Tokyo 181-8588, Japan

T. Iwata, T. Yokoyama
National Space Development Agency of Japan, 2-1-1 Sengen, Tsukuba, Ibaraki 305-8505, Japan

H. Kanamori
Institute of Technology , Space Systems Laboratory,
Shimizu Corporation, 3-4-17, Etchujima, Koto-ku, Tokyo,135-8530, Japan

K. Funazaki,
Faculty of Engineering, Iwate University, 4-3-5 Ueda, Morioka-shi, Iwate 020-8551, Japan

T. Miyazaki
Faculty of Engineering,
Kagoshima University, 1-21-40 Korimoto, Kagoshima-shi, Kagoshima 890-0065, Japan

Abstract. We are proposing a selenodetic mission, e.g. In situ Lunar Orientation Measurement (ILOM) to study lunar rotational dynamics by direct observations of the lunar physical libration and the free librations from the lunar surface with an accuracy of 1 millisecond of arc in the post-SELENE project. Year-long trajectories of the stars provide information on various components of the physical librations and they can also be used to possibly detect the lunar free librations in order to investigate the lunar mantle and the liquid core. The PZT on the moon is similar to that used for latitude observations of the Earth except the half mirror above the objective, a CCD with high well capacity, and the heater beneath the mercury pool. Although a star position on the focal plane does not change even if the telescope inclines in principle, the tilt of the telescope affects the star position due to aberrations of the objective in the actual case. We obtained the relation between the deviation of the star position in CCD and the tilt of the telescope by ray tracing method and found that the effect of the tilt less than 100 arc seconds upon the star position does not exceed 1 mas. Thermal test of a tube made of CFRP showed that the tube did not incline by more than 100 seconds of arc even in the conditions of the lunar surface. We have a prospect to attain an accuracy of positioning of better than 1 mas from simulated experiments in laboratory using a CCD.

Keywords. Physical librations, photographic zenith tube (PZT), lunar core

1 Introduction

Whether the Moon has a metallic core or not and whether it is melted or not are important problems related to origin and evolution of the Moon and the solar system. Structure and property of the interior is estimated mainly by array observations of earthquakes and observations of rotation in the case of the Earth. We know how the mass is

concentrated to the center from the moment of inertia which is estimated by rotational parameters and low degree spherical harmonics of gravitational fields. The seismic observations, on the other hand, can detect boundaries of the core, the mantle and the crust through travel-time of seismic waves. The property of the core, which is more difficult to observe, is estimated by how the waves propagate in the core and by dissipation and resonance of rotation and tides.

The moon has rotational fluctuations like the Earth which are called by the name of physical and free librations. The physical librations are caused by an attraction between the moon and the planets. The free librations, on the other hand, were considered to be caused by impacts of meteorites or asteroids in early days and it was later shown by Peale (1976) that impacts are not frequent enough to excite sizable free librations. Among many kinds of observations of the Moon, the physical librations have a potential to bring about important information especially on deep interior.

Amplitudes of the physical librations combined with low degree spherical harmonics of the lunar gravity fields can give a constraint upon density of the lunar core through the moment of inertia. The moment of inertia of the Moon was estimated at 0.3929+/-0.0009 in 1990's and it was suggested that the Moon has a small metallic core since the value is slightly smaller than 0.4 which means a homogeneous Moon (Williams et al., 1996). This result, however, could not clearly show the existence of the metallic core due to the ambiguity in the value itself and the insufficient evaluation of the effect of upper layer such as the crust and the mantle (Dickey et al. (1994)).

The accuracy of the moment of inertia was improved to 0.3932+/-0.0002 after the Lunar Prospector mission and the radius of the core was estimated as 320km+50/-100km if the material was supposed to be Fe and 510km+80/-180km if it was FeS (Konopliv et al. (1998)). The range of radius which the core can have was first shown from selenodesy data combined with the librations and the gravity fields. More accurate value of the moment of inertia is necessary for this study since the sensitivity to density of the core which is concentrated on the center is very weak and since the effect of inhomogeneity in the crust and difference of the moments of inertia with respect to different axes are in the order of 0.1 % of the moment of inertia.

Williams et al. (2001) recently detected phase differences due to dissipation inside of the Moon in various components of the physical librations from 30 year LLR data and suggested dissipation within the core-mantle boundary which means that the Moon has a liquid core. We can get new information related to properties of the lunar core from improved accuracy of forced physical librations as well as the amplitudes and periods of the free librations (Newhall and Williams, 1997). Another type of observation with higher accuracy will contribute to conclude this important problem.

Seismic data, on the other hand, do not give enough information of velocity structure of deep interior since the network of seismometers on the Moon is restricted to the near side of the Moon.

We are proposing a selenodetic mission, In-situ Lunar Orientation Measurement (ILOM) to study lunar rotational dynamics by direct observations of the lunar physical libration from the lunar surface by using a photographic zenith tube (PZT) as a candidate for the post SELENE project. Outline of the project, expected accuracy and technical problems are presented in this paper.

2 Roles of a New Observation

Lunar Laser Ranging (LLR) observations for about 30 years are being done and they have contributed to improving the lunar libration models including the free librations (Dickey et al., 1994). LLR can determine parameters of the physical librations with an relative accuracy of about 10^{-4} from the ranging accuracy of a few centimeters. However, it is not always enough to estimate the libration parameters with small amplitudes such as the free librations and the dissipation terms from the LLR data. We must consider the following points when we evaluate LLR observations.

1) Observations of the physical librations from the Earth such as LLR are affected by the Earth rotation and the optical librations with larger amplitudes,
2) LLR has no sensitivity in the direction perpendicular to the Earth-moon direction and therefore it has less sensitivity for observations of the libration around the axis of the Earth-moon direction,
3) LLR observations are difficult during periods of new and full moons and observation period is synchronous with a libration period although it is not a major term,
4) Amplitudes of the physical librations are affected by uncertainties in the position of the corner reflectors.

It can not be excluded that there still remain systematic errors in the libration parameters determined from LLR data. A new kind of observation which is independent of LLR and having higher accuracy will conclude these problems. The new observation should not only improve accuracy but also help separation of different parameters and separation of different frequencies in order to solve the problems related to

the deep interior of the Moon by the co-operation with improvement of LLR accuracy (Murphy et al., 2002).

3 In situ Lunar Orientation Measurement (ILOM) project

In the ILOM project, a photographic zenith tube (PZT) set near the lunar pole determines the orientation of the axis of rotation of the Moon by positioning of several tens of stars in the field of view at every moment for longer than one year. Variation of spin rate which is the third component of the librations and is insensitive to the observation of stars in the polar region is also measured by putting a slanting half mirror on the top of the telescope for stars to be observed near the equator although the positioning accuracy is directly affected by deformation or displacement of the mirror. It is necessary to develop a very stable and precise holding mechanism for the mirror in order to achieve the same accuracy as that for polar stars.

The PZT is suitable for a positioning telescope on the moon since a mercury pool set at the middle point of the focal length compensates the tilt of the telescope caused by unequal thermal expansion. Star positions are recorded with a CCD camera with 4,000×4,000 pixels set beneath an objective for an integration time and the center of each star image at a certain time is determined by modeling a point spread function of the image including the effect of defocusing and lunar motion.

The accuracy of ILOM should be higher by about one order of magnitude than that of LLR or than 1 millisecond of arc even if it is a new kind of observation since libration parameters related to property of the lunar core have amplitude of at most a few milliseconds of arc (Dickey et al., 1994). Although a longer focal length gives rise to more accurate observation, we assume to adopt a focal length of 2m considering limitations of a payload. Then the resolution of the telescope or the diffraction limit is 1 second of arc for wavelength of 500 nm.

The resolution of 1 second of arc corresponds to about 10μm on the focal plane. The minimum CCD pixel, on the other hand, has a dimension of about 10μm×10 μm by chance. We defocus a star image and determine the center of the image spreading over several pixels by fitting to a Gaussian distribution model in order to improve the apparent resolution. Simulations show that the positioning error in the CCD plane is less than 1/1,000 of a pixel size in the condition that an exposure time is 100s, a pixel size is 9 μm, magnitude of the star is brighter than 12 and the star image distributes over a few pixels squares (Hanada et al., 2000). Specifications of the PZT for ILOM project are listed in Table 1 and schematics of the telescope is shown in Fig. 1.

Table 1. Specifications of the PZT

Aperture	0.2m
Focal Length	2m
Resolving Power	1" (λ=500nm)
Detector	CCD
Dimension of a Pixel	9μm×9μm
Number of Pixels in an Array	4,096× 4,096
Field of View	1°×1°
Integration Time	10s
Magnitude of Target Stars	M < 12

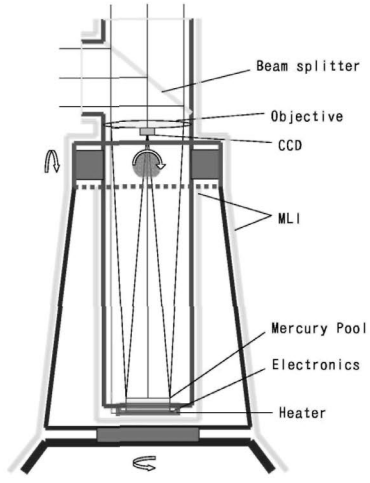


Fig. 1. A schematic diagram of the PZT on the moon

4 Evaluation of ILOM

There are advantages in ILOM which aims at observations of the libration parameters with less than 1 millisecond of arc by optical astronomical observations from the lunar surface,

1) ILOM has sensitivity in the direction perpendicular to those of LLR,
2) ILOM is independent on the effects of the rotation of the Earth and the lunar ephemeris,
3) Simulation shows that ILOM can determine the libration parameters with about one order higher than LLR by positioning of stars with the accuracy of 1 millisecond of arc (Heki et al., 2000),
4) Simulation shows that ILOM can distinguish the free libration terms from the other physical libration terms by observations for longer than one year (Heki et al., 2000),

5) Determination of rotation parameters by optical astronomical observations has given many results in the study of Earth's rotation.

Table 2. Comparison with other method

	Independency	Accuracy	Remark
ILOM	Independent	High	1)
Laser Ranging on the Moon	Dependent	High	2)
VLBI、GPS	Independent	High	
Doppler	Dependent	Low	3)
Gravimeter/Tiltmeter on the Moon	Independent	Low	4)

1) Observation on the polar region is effective, 2) Independent to LLR if observations are made between the lunar surface and an orbiter, 3) Less sensitive to the rotation, 4) Less sensitive to the rotation and unstable for long term

Table 3. Evaluation of the lunar surface

	Moon	Satellite	Earth
Flicker of the Air	No	No	Affected[1)]
Vibration Noise	Quiet	Noisy[2)]	Noisy[3)]
Stable base	Available[4)]	No[5)]	Available
Temperature Condition	Severe[6)]	Severe[6)]	Mild
Dust	Dusty[7)]	Clean	Clean
Mechanical Control	Difficult[8)]	Difficult[8)]	Easy
Maintenance	Difficult	Difficult	Easy

1) limits the accuracy, 2) Reaction wheel, 3) Continual microseisms, 4) Tsubokawa (2003), 5) Attitude control, 6) Thermal control is necessary, 7) Protection is necessary, 8) Friction in a vacuum

VLBI or SLR replaced the optical observations in the case of the Earth in order to overcome the limit of accuracy which comes from a flicker of the air. There seems to be no factor which limits the accuracy for the optical observations on the Moon. Further, there is no restriction on observation time if the telescope is set in the polar region.

Different kinds of observations can be used to investigate the interior of the Moon as in the case of the Earth. We can evaluate the methods from the points of view of 1) independency on LLR and 2) potential accuracy relative to LLR as shown in Table 2. Effectiveness of ILOM is obvious from Table 2.

We evaluate the observation condition on the lunar surface in order to show that there is no factor which limits the accuracy. Characteristics of the Lunar surface as an observation station are listed in Table 3 by comparing with the ground and an artificial satellite.

The effects of flicker of the air and the seismic noise are fatal for precise positioning and they limit the accuracy in observation on the ground although the ground seems to be a good place for observations in practical point of view. The lunar surface, on the other hand, is an ideal place keeping a possibility of an epoch-making observation if we overcome the less important problems such as a big temperature change and the dust originating from the regolith.

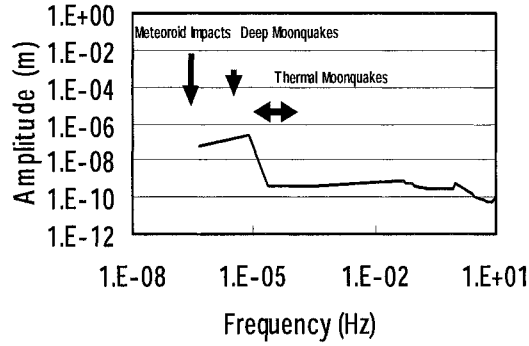


Fig. 2. Seismic noise level on the moon

Table 4. Error sources

Error sources	Components
Optical path length	Expansion of a tube
	Deformation of an objective
	Tilt of an optical parallel
Deflection	Tilt of the tube
	Tilt of the objective
	Unsymmetrical deformation of the objective and mercury surface
Reading error	Misfit to a point spread function
	Seismic noise
Level surface	Tidal change
	Gravity anomaly

It is needless to say that the lunar surface is in an ideal condition from the point of view of flicker noise since there is no atmosphere affecting the observation. We can also show that the seismic noise level on the lunar surface is very low.

Figure 2 shows seismic noise on the Moon, inferred from data for periods before moonquakes occur, recorded by long period seismometer set by the Apollo project. The amplitude of the vibration is less than that which corresponds to 1 bit of A/D converter and we suppose here that the maximum amplitude during a quiet time is the amplitude corresponding to 1 bit. Maximum amplitudes and frequencies of moonquakes caused by various reasons such as impacts, thermal stress and tidal stress, or deep moonquakes, are also shown as

comparison. This level is two or three orders of magnitude smaller than that of the Earth (Hanada et al., 2003).

5 Technical Problems in ILOM

There are many technical problems in ILOM project in order to attain the accuracy of 1 millisecond of arc and in order to realize the observation on the Moon in the first place. Error sources which may affect the accuracy are divided into four groups: 1) change of optical path length, 2) deflection of a ray, 3) reading error, and 4) change of a level surface as a reference frame. The error sources are traced to components or environment of the telescope as shown in Table 4.

In principle, a star position on the focal plane does not change even if the telescope inclines a little. However, the tilt of the telescope affects the star position due to aberrations of the objective in actual case. We investigated the relation between deviation of the center of a star image in CCD and a tilt of the telescope by a ray tracing method. Table 5 shows deviations of center of gravity, which means a mean position of a lot of rays, 1) on the focal plane (upper), 2) on the plane 0.6mm farther than the focal plane, 3) on the plane 0.6mm nearer than the focal plane, taking incident angles and tilt of the tube as parameters. An objective with two convex lenses including aspherical surfaces is used here.

Table 5. Deflection of the center of star images

1) On the Focal plane				
		Incident angle		
		0 deg	0.25 deg	0.50 deg
Tilt	0deg	0.0 mas	7.5 mas	14.5 mas
	+0.1deg	−3.1 mas	4.5 mas	11.8 mas
	-0.1deg	+3.1 mas	10.4 mas	17.2 mas

2) +0.6 mm farther than the focal plane				
		Incident angle		
		0 deg	0.25 deg	0.50 deg
Tilt	0deg	0.0 mas	7.2 mas	14.0 mas
	+0.1deg	-2.8 mas	4.4 mas	11.4 mas
	-0.1deg	2.8 mas	9.9 mas	16.4 mas

3) +0.6 mm nearer than the focal plane				
		Incident angle		
		0 deg	0.25 deg	0.50 deg
Tilt	0deg	0.0 mas	7.7 mas	15.1 mas
	+0.1deg	-3.3 mas	4.5 mas	12.1 mas
	-0.1deg	3.3 mas	10.9 mas	17.9 mas

mas: millisecond of arc

The effect of inclination of the tube upon the deviation of star images is estimated at about 3 milliseconds of arc (3mas) with the tilt of 0.1 degrees. This means that the tilt of the tube must be kept within about 100 seconds of arc for the measurements of 1mas accuracy. Tsuruta et al. (2003) experimentally showed that the tilt did not exceed 100 seconds of arc even under the condition of temperature difference between opposite sides by 50 degrees by using a thermal model of a tube made of CFRP (Carbon Fiber Reinforced Plastic) surrounded by several layers of thermal insulators. To realize the condition where the temperature difference is lower than 50 degrees on the lunar surface is a future problem.

Tsubokawa (2003) made experiments from another point of view of characteristics of the lunar soil called regolith, and showed that a base on the lunar soil which supports the telescope did not incline by more than 1 second of arc due to unbalanced load.

The deviation exceeds 1 mas when an incident angle becomes larger than about 130 seconds of arc even if the tube does not incline. It is not critical because we can avoid this effect by correcting the effect using the relation between the deviation and the incident angle. It is also possible to minimize the effect by adjusting parameters of the objective.

Vertical position of CCD is not very much sensitive to the deviation. The position change of star images is small in the range of 0.6 mm apart from the focal plane. Requirement coming from this result is relatively loose and a tube of 1m long will not expand by 0.6 mm if the temperature does not change by more than 600 degrees.

Centroid experiment with a CCD camera was initiated aiming at precise positioning of simulated star images on fine controlled CCD camera in cooperation with infrared space astrometry project of NAOJ (JASMINE). It has been confirmed that the positioning of pseudo star is possible with the accuracy of 1/300 pixel using the photon weighting mean method. This is the same level record as FAME group had achieved. Centroid experiments and analysis will be continue for the ILOM goal of 1/1000 pixel accuracy considering higher order approximation for the difference between the true center and the measured center of a star image (Araki et al., 2003).

The effect of tidal tilts and gravity anomaly upon the level surface are essential but can not be attributed to the instrument itself. The effect of the tides which has amplitude far larger than 1 mas and has the same periods as the physical librations must be removed. Accurate values of the Love numbers and accurate position of the telescope are necessary for the correction satisfying the requirement of

ILOM.

Analysis of deformation of optical elements and development of controlling the effects will be concluded in the near future.

6 Concluding Remarks

We introduced a selenodetic mission, the ILOM project, which observes the lunar physical librations by a PZT telescope on the lunar surface, and discussed feasibility, expected scientific output, error budget and technical problems as compared with LLR. It is concluded that:

1) observations of the lunar physical librations by new observations with the accuracy one order higher than before can derive important information of the lunar core related to the origin and evolution of the Moon,
2) astronomical positioning on the Moon is effective as the new observation,
3) it is important to develop technologies for controlling the lunar environment in order to realize the ILOM project.

Important scientific discussion remained after the investigation so far are:

1) how accurately the ILOM project can measure the third angle, the rotation about the pole,
2) how to separate the influence of the tides,
3) how long we should observe on the Moon in order to separate important libration terms with a sufficient accuracy.

Remaining technical investigations are:

1) analysis of deformation of the mercury surface and the objective due to temperature change,
2) development of a method to control the environment such as temperature and dust,
3) development of a software to determine lunar orientation parameters from star image data.

Acknowledgements

The authors wish to express their gratitude to Drs Jim Williams and Daniel Gambis for their critical reviews and valuable comments.

References

Araki, H., S. Tazawa, K. Asari, T. Tsubokawa, S. Tsuruta, H. Hanada, N. Kawano, T. Yano, N. Gouda and RISE Research Group. (2003). Centroiding Experiment of Artificial Stars. *Annual Report of NAOJ*, (in press).

Dickey, J. O., P. L. Bender, J. E. Faller, X. X. Newhall, R. L. Ricklefs, J. G. Ries, P. J. Shelus, C. Veillet, A. L. Whipple, J. R. Wiant, J. G. Williams and C. F. Yoder (1994). Lunar laser ranging: a continuing legacy of the Apollo program. *Science*, 265, pp. 482-490.

Hanada, H., K. Heki, N. Kawano, M. Ooe, T. Tsubokawa, S. Tsuruta, T. Ishikawa, H. Araki, K. Matsumoto, T. Takanezawa, Y. Kono, H. Karoji, T. Iwata, Y. Kaneko and T. Yokoyama (2000). Advanced Observations of Lunar Physical Librations and Gravitational Fields in Japanese Lunar Missions in the Near Future. *ASP Conference Series*, 208, pp. 623-630.

Hanada, H. and ILOM Research Group (2003). Observation of Lunar rotation on the Moon. *Proc. Geodesy beyond the Earth*, 40-44 (in Japanese).

Heki, K.. (2000). Observation of the lunar physical libration and tidal deformation by ILOM (In-situ Lunar Orientation Measurement). *Proc. Symp. on Tidal Studies in Tectonic Active Regions*, 99-104 (in Japanese).

Konopliv, A.S., A.B. Binder, L.L. Hood, A.B. Kucinskas, W.L. Sjogren and J.G. Williams (1998). Improved gravity field of the Moon from Lunar Prospector. *Science*, 281, pp. 1476-1480.

Murphy, T.M., J.D. Strasburg, C.W. Stubbs, E.G. Adelberger, J. Angle, K. Nordtvedt, J.G. Williams, J.O. Dickey and B. Gillspie (2002, The Apache Point Observatory Lunar Laser-Ranging Operation (APOLLO), Proc. 12th Int. Workshop on Lawer Ranging, Matera, Italy (in press).

Peale, S.J. (1976), Excitation and relaxation of the wobble, precession, and libration of the Moon, *J. Geophys. Res.*, 81, 1813-1827.

Newhall, X.X. and J.G. Williams (1997), Estimation of the lunar physical librations, *Celestial. Mech. Dynamical Astron.*, 66, 21-30.

Tsubokawa, T. (2003). Experiments on stability of the lunar soil. *Proc. Geodesy beyond the Earth*, 116-125 (in Japanese).

Tsuruta, S., T. Nomiyama, Y. Horie, H. Hanada, N. Kawano and T. Miyazaki (2003). Thermal Expansion coefficient of a telescope tube prepared for observations on the Moon. *Proc. Nat. Astr. Obs. Japan*, (submitted)

Williams, J.G., X.X. Newhall and J.O. Dickey (1996). Lunar moments, tides, orientation, and coordinate frames. *Planet. Space Sci.*, 44, pp. 1077-1080.

Williams, J.G., D.H. Boggs, C.F. Yoder, J.T. Ratcliff and J.O. Dickey (2001). Lunar rotational dissipation in solid body and molten core. *J. Geophys. Res.*, 106, pp. 27933-27968.

Combined Processing of GPS Data from Ground Station and LEO Receivers in a Global Solution

U. Hugentobler, A. Jäggi, S. Schaer, G. Beutler
Astronomical Institute, University of Berne, Sidlerstrasse 5, CH-3012 Berne, Switzerland
e-mail: urs.hugentobler@aiub.unibe.ch

Abstract. GPS plays an important role for the definition of the terrestrial reference frame and for the monitoring of Earth orientation parameters based on observations acquired by the global IGS (International GPS Service) tracking network. With an increasing number of low Earth orbiters (LEOs) carrying high-quality dual-frequency GPS receivers the question arises whether the inclusion of orbiter receiver data may help to further improve parameters estimated in a global GPS data analysis.

In the following study the satellite JASON-1 was introduced a posteriori into the global analysis performed at the CODE Analysis Center of the IGS. A relatively simple strategy was used allowing for a double difference data analysis applying correct correlations and taking advantage of the fact that preprocessed files with fixed ambiguities were already available for the ground network.

An improvement of the LEO orbit as a result from the combined processing could not be clearly shown. The impact on GPS orbits and geocenter coordinates caused by the inclusion of JASON is, on the other hand, significantly larger than expected from the addition of one receiver to a 120 station tracking network. This indicates a potential for improvements with respect to these parameters by means of a combined processing. Modeling insufficiencies and inconsistencies, however, need to be solved to reach this goal.

Key words. Low Earth orbiters – Orbit determination – Single-step approach – Two-step approach

1 Introduction

The GPS (Global Positioning System) plays an important role for the definition of the terrestrial reference frame and for the monitoring of Earth orientation parameters. This work is performed by IGS Analysis Centers which use tracking data acquired by the global station network of the IGS. On the other hand, more and more low Earth orbiters carrying high-quality GPS receivers are or will soon be available. With the upcoming satellite constellations for atmospheric sounding such as COSMIC (Rocken et al., 2000) soon more than a dozen LEOs with on-board geodetic-type GPS receivers will orbit the Earth. These satellites may be considered as additional stations with different properties complementing the ground tracking IGS network.

It is an unanswered question whether the inclusion of LEOs into the analysis of global GPS data helps to improve the accuracy of the estimated Earth orientation parameters, geocenter coordinates, or GPS orbits. It is yet unclear, as well, by how much the accuracy of the orbit of a LEO may be improved by solving for the orbit parameters together with the global parameters (GPS orbits, Earth orientation parameters, station coordinates, troposphere parameters, etc.; so-called *single-step approach*) as compared to a procedure with the global parameters introduced as fixed (*two-step approach*).

Already nearly a decade ago Rim et al. (1995) performed tests with single-step and two-step solutions of GPS-derived precise orbits for TOPEX/POSEIDON. A minor improvement in the radial and along-track component for the LEO orbit was claimed. Tests carried out by Zhu et al. (2002) using observations stemming from the CHAMP satellite indicate an improvement of the reference frame realization as well as of the quality of the LEO and GPS orbits. Based on simulations assuming optimum conditions Hugentobler and Beutler (2003) found a small positive impact on the formal accuracy of along-track and cross-track positions for GPS satellites.

In a study on precise orbit determination for the GOCE satellite, on the other hand, Visser and van den IJssel (2001) did not find a significant improvement of either the LEO or the GPS orbits. It may, indeed, be pointed out that, in order to expect an improvement of the parameters of a global solution, the dynamic or reduced-dynamic orbit for the LEO has to be accurate at the centimeter level.

To approach the answer to the above questions GPS tracking data from JASON-1 covering six days (June 3 to 8, 2002) overlapping with the JASON comparison campaign of the IGS LEO Pilot Project (Boomkamp, 2002) were used. During this time period the JASON GPS receiver was tracking eight satellites on the average. The six days studied con-

tain 25 data gaps with a total length of 190 minutes (about 2% of total tracking time).

The JASON data were processed in the double-difference mode using tracking data from the IGS network. In all cases reduced-dynamic orbital arcs covering 24 hours were computed with stochastic pulses (velocity changes) introduced every 12 minutes. The phase center position of the JASON GPS antenna with respect to the satellite's center of mass was taken from Haines et al. (2002). For all computations a development version of the Bernese GPS Software Version 5.0 (Hugentobler et al., 2001) was used.

2 Final Processing at CODE

To realize the single-step approach the JASON data were introduced into the processing chain used at CODE (Center for Orbit Determination in Europe, (Hugentobler et al., 2003)) to compute the final products for the IGS. Before outlining the combination strategy we briefly describe this processing procedure.

Before increasing the maximum number of processed IGS stations to 150 in July 2002 typically 120 tracking stations were included in the processing. Precise GPS orbits, station coordinates, Earth orientation parameters, troposphere and ionosphere parameters, differential code biases, and satellite and receiver clock corrections, are computed on a daily basis. The processing is based on double differences of the ionosphere-free linear combination of the two carriers L_1 and L_2 (except for the fixing of ambiguities, the computation of clock corrections, and the determination of ionosphere parameters). Carrier phase ambiguities are solved in several steps for baselines with lengths of up to 6000 km. An elevation cut-off angle of 3 degrees is used. Twelve zenith path delay parameters as well as one pair of tropospheric gradient parameters are estimated per day for each site.

Four regional clusters of baselines are processed in parallel and independently of each other in order to increase the efficiency. The results from all clusters are then combined on the normal equation level. To generate the final products delivered to the IGS the normal equations of successive three days are accumulated to generate a three-day arc for the GPS satellites. Seven of the resulting normal equations are combined in order to compute weekly station coordinates which, in turn, are back-substituted into the three-day normal equations to further improve GPS orbits and troposphere parameters. The final orbits of a day delivered to IGS represent the middle day of a three-day orbit.

For our study no computations exceeding the one-day level were performed, i.e., all computed satellite orbits are one-day arcs. The study aimed at the inclusion of baselines space–ground into the individual clusters of ground baselines. Using archived single difference files from the CODE final processing, advantage could be taken of the fact that for the ground network about 85% of the carrier phase ambiguities were already fixed to integer numbers. No attempt was made to solve the ambiguities for the baselines relating the ground stations to the LEO.

3 Combination Strategy

The baselines space–ground were generated in such a way that exactly one of them was active for a given epoch. The generated single-difference files can then easily be added to the clusters of ground baselines. An alternative strategy would allow for more than one baseline space–ground that is active at a time. Epochs at which one of the baselines is replaced by another baseline due to the motion of the satellite could then be 'bridged' by the other baselines which would stabilize the solution. In this case, however, baselines in the ground network would have to be temporarily removed in order not to generate redundant (linearly dependent) double difference observations.

Candidate stations for forming baselines space–ground were selected based on two conditions: (1) They have to track the same satellites as JASON. One missing satellite at maximum was allowed. (2) For the final set of baselines those stations were selected, which allowed for a maximum time interval during which the baseline could be active. Based on these two conditions an automatic procedure was implemented which generates baselines space–ground with a minimum number of switches from one ground station to the next and with a maximum number of GPS satellites observed by both receivers.

With the outlined procedure typically 70 baselines per day were formed between JASON and a ground receiver including about 40 ground stations. Isolated sites were selected several times per day. Baselines were 'active' between 4 and 40 minutes. On average baselines were active for 20 minutes before being replaced by another, i.e., one orbital revolution of the satellite was covered by 4–5 baselines on average. A total number of 70 stations out of the total of 120 stations were selected as endpoints of baselines space–ground during the six days considered. These stations are marked in Figure 1 with circles with diameters proportional to the number of passes found. Large circles, prominent for isolated sites, indicate the importance of a station for the tracking of a LEO. For Rio Grande (RIOG, Argentina) a total of 17 passes were formed within the six day period, followed by Kourou (KOUR, French Guyana) with 15 passes.

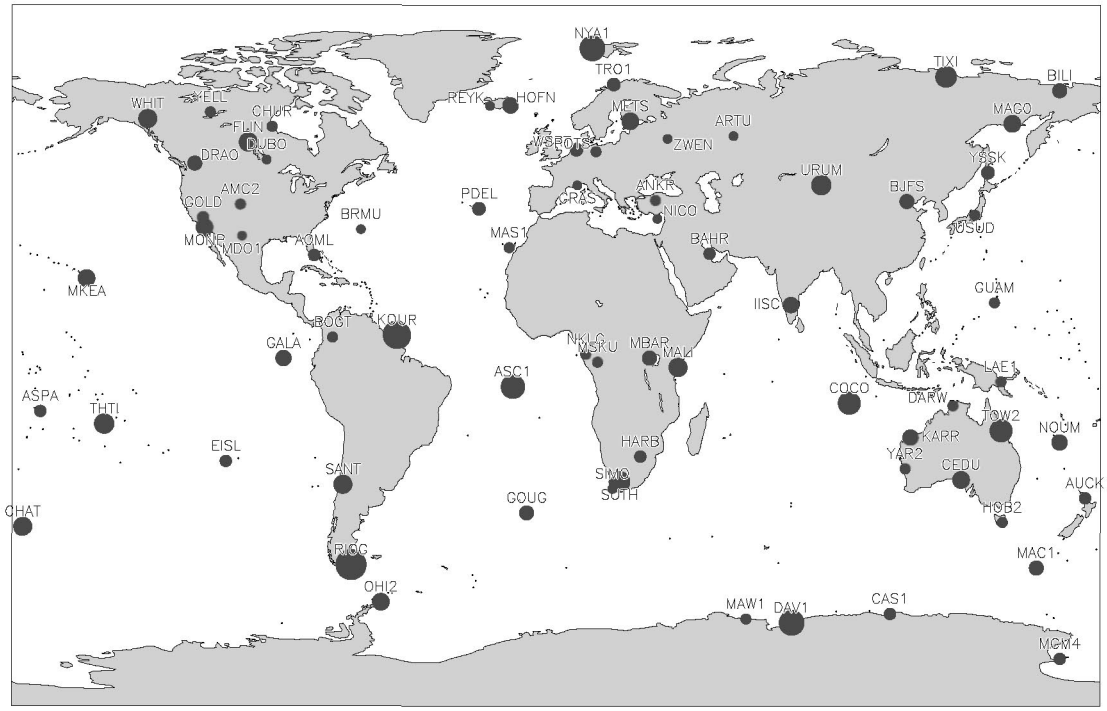


Fig. 1. Selected GPS tracking sites to form baselines with JASON for a time period of six days. The diameter of the circles is proportional to the number of passes.

4 Characteristics of Solutions

Table 1 contains statistical information for the solutions generated for June 3, 2002. The first column gives the numbers for the processing of the global network including GPS orbits but without inclusion of JASON, i.e., the numbers for the Final processing at CODE at the one-day level (first step of the two-step approach). The second column gives the corresponding numbers for the computation of the LEO orbit with all information from the global network fixed, i.e., GPS orbits, Earth orientation, station coordinates, and troposphere parameters (second step of the two-step approach). The last column gives the numbers for the solution containing all parameters, including the LEO orbital parameters (single-step approach).

The orbits of both, GPS satellites and JASON, are modeled by six initial conditions, nine radiation pressure parameters according to the CODE extended radiation pressure model (Beutler et al., 1994), and stochastic pulses. For the GPS satellites a single set of constrained pulses in three orthogonal directions are set up at noon, for JASON stochastic parameters are set up every 12 minutes in order to account for possible deficiencies of the orbit model used for the low orbiting satellite.

Tropospheric refraction for ground stations is modeled by two-hourly zenith path delay corrections and one set of tropospheric gradient parameters per day. The cut-off elevation angles are set to 3° for ground stations and 0° for JASON. For ground stations elevation dependent weighting proportional $\cos^2 z$ was enabled (z is the zenith angle). Sampling rates are 180 seconds for ground stations and 30 seconds for the LEO. In the case of the single-step approach the higher sampling rate is applied to all baselines containing the satellite.

Comparing the final processing without (first column in Table 1) and with LEO (last column) one recognizes a significant increase of the number of parameters. The explicit number of parameters which give the dimension of the normal equation matrix which has to be inverted is increased by 372 parameters, most of them stochastic orbit parameters. The baselines space–ground add 800 to 1000 ambiguities to the unresolved float ambiguities at the ground tracking network.

Table 1. Statistical information of the solutions for doy 154/2002

Solution	CODE Final	LEO only	GPS+LEO
Number of stations	119	–	119
Number of single difference files	142	43	214
Number of satellites	28	1	29
Number of observations	390908	13268	414569
Sampling rate (seconds)	180	30	180/30
Number of explicit parameters	2896	372	3268
Coordinates	357	–	357
Troposphere	2022	–	2022
Orbital parameters	420	15	435
Stochastic orbit parameters	84	357	441
Earth orientation parameters	10	–	10
Total number of parameters	6411	910	7756
Unresolved ambiguities (pre-eliminated)	3515	538	4488

5 Validation of LEO Orbit

To validate the LEO orbits resulting from the single-step and the two-step approach, several criteria were studied: overlap at the arc boundaries, SLR (Satellite Laser Ranging) residuals, and differences to external JASON orbits generated by CNES (Centre National des Etudes Spaciales) based on SLR and DORIS (Doppler Orbitography by Radiopositioning Integrated on Satellite) observations and delivered to the IGS LEO Pilot Project. Figure 2 shows these quality indicators as a function of the constraints applied to the stochastic pulses for JASON orbits originating from the single-step approach for the first of the six days processed (154/2002). The left hand side part of the figure (tight constraints) represents the dynamical limit of the LEO orbit while the right hand side part (loose constraints) approaches the kinematic limit.

The 3-D position differences of the orbit overlap with the following arc show a pronounced dependence on the constraints. For tight constraints the overlap agreement increases to 18 cm, revealing deficiencies of the dynamic orbit model used for JASON. For too loose constraints some of the stochastic parameters get singular occasionally which degrades the solution. For the best solution a 3-D overlap of 3.5 cm is obtained. The overlap in radial direction then amounts to 8 mm. The RMS of differences per coordinate to the orbit from CNES shows an minimum value of 3.2 cm. Values for the other days reach 5.7 cm. No significant differences in orientation and scale are found, but translations between the two orbits may reach up to 3 cm for the time interval considered which indicates small inconsistencies between the solutions.

The SLR residuals for the processed day with the largest number of SLR normal points (170 observations for doy 154/2002) show an RMS at the level of 2.7 cm. The top panel of Figure 3 shows the RMS of

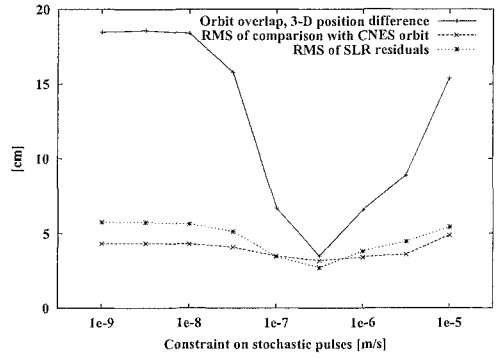


Fig. 2. Variation of constraints on stochastic orbit parameters for JASON for doy 154/2002. Influence on orbit overlaps, on differences to external orbit, and on SLR residuals.

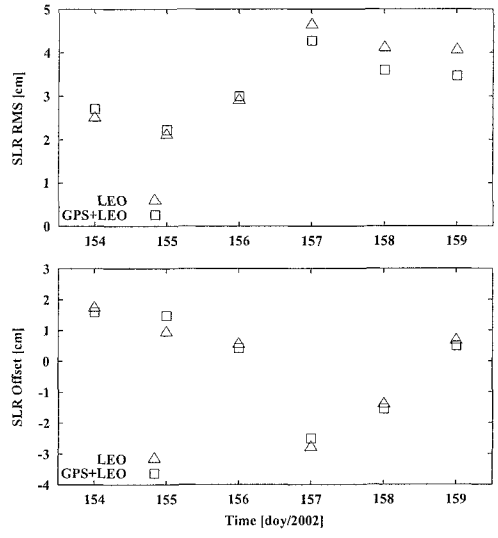


Fig. 3. SLR RMS values of residuals (top), and SLR offset (bottom), for all processed days based on two-step approach (triangles) and single-step approach (boxes).

the SLR residuals for all computed days, the bottom panel of the same figure gives the corresponding values of the range biases (included in the RMS given above). Since no SLR bias is expected a value differing from zero indicates some inconsistencies in the solutions. The figures give the values for the two-step approach (triangles) and for the single-step approach (boxes). The average SLR RMS amounts to 2.4 cm for the single-step and 3.0 cm for the two-step approach. The mean bias is 0.5 cm and 0.8 cm respectively. The SLR RMS does indicate a small improvement of the LEO orbit for the single-step approach when compared to the two-step approach. The number of solutions is, however, too small to give a firm answer on whether the single-step gives superior results for the LEO orbit.

6 Impact on Global GPS Solution

Including JASON into the global network affects the GPS satellite orbits at the centimeter level. The top panel in Figure 4 shows the RMS difference per coordinate between GPS orbits computed without and with inclusion of JASON. The RMS values increase from about 0.5 cm to 1 cm when passing from a reduced-dynamic LEO orbit (loose constraints) to a nearly dynamic LEO orbit (tight constraints) indicating the impact on the GPS orbits.

No significant change in scale and orientation of the GPS orbits is found when adding JASON to the processing. However, as can be seen in the center panel in Figure 4, a translation of the constellation of a few millimeters is caused by adding the LEO. Obviously the 'geocenter' is sensitive to the inclusion of low Earth orbiting satellites.

The agreement of the station coordinates computed with and without JASON is at the submillimeter level after applying a Helmert transformation. No significant scale is found. Because the datum is defined in both cases using a no-net-rotation condition with respect to IGS00 (Ferland, 2002), rotations are not expected. But again, translations between the two network solutions are significant, in particular in the z-direction. The translation in z is of the order of 1 cm and shows a larger scatter from day to day. It is interesting to note that introducing an inconsistency intentionally by changing the JASON antenna phase center position particularly affects the z-translation of the GPS constellation as well as of the ground network.

The Earth orientation parameters show differences of the order of a few ten μas for pole coordinates, up to 0.2 mas/day for the pole rates, and up to 10 μs/day for length of day.

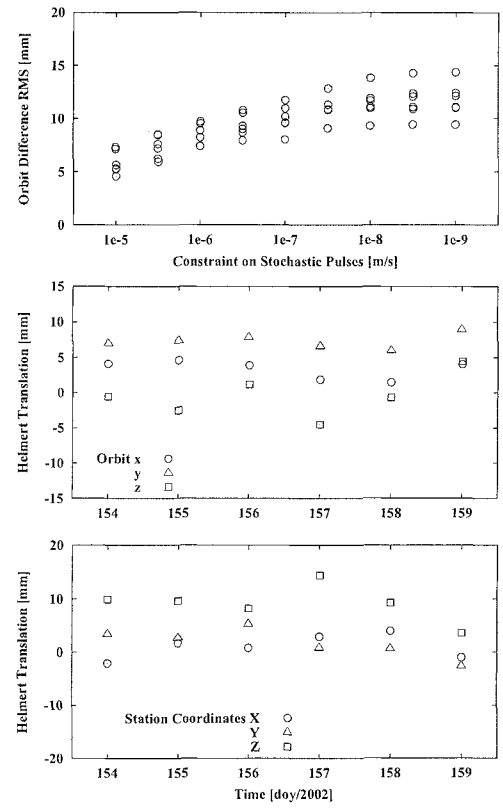


Fig. 4. RMS of differences per coordinate between GPS orbits computed with and without JASON as a function of the stochastic constraints for the LEO orbit for the six days processed (top), translation of the two GPS orbit solutions (center), and of the station coordinate solutions (bottom).

7 Summary and Conclusions

Using a simple combination strategy double difference observations on baselines space–ground could be added to preprocessed baselines of the IGS subnetwork of GPS tracking sites processed at the CODE Analysis Center. The procedure allows to include LEOs into the CODE global data analysis based on 120 ground stations and to take maximum profit of already fixed ambiguities in the ground network. The strategy applied to JASON GPS tracking data is simple, but allows for a processing of the global tracking network together with baselines space–ground with correct mathematical correlations.

The quality of the reduced-dynamic JASON orbits is of the order of 3-5 cm as confirmed by internal (orbit overlap) and external comparisons (SLR as well as orbits provided by CNES derived from SLR and DORIS observations). The comparisons with the CNES orbit show significant translations of up to 3 cm. A plain comparison between the orbits (no

Helmert transformation applied) results in a RMS difference per coordinate of 3-5 cm. The SLR residuals confirm these numbers. RMS values range between 2 cm and 5 cm.

An improvement of the LEO orbit by applying a single-step approach instead of a two-step approach could not be convincingly demonstrated by the limited experiment based on 6 days of JASON data. Results from more days have to be available to favor one of the two approaches.

Concerning the parameters of the global solution an impact of the order of one centimeter on the GPS orbits could be observed when adding JASON into the system. Modeling inconsistencies in conjunction with the inhomogeneous geographical distribution of the GPS tracking sites might cause the effect. For the station coordinates a translation of about 1 cm could be observed in the direction of the Earth's rotation axis. The impact on the station coordinates themselves, however, resides at the sub-millimeter level. For Earth orientation parameters the largest impact was found for the daily pole drift rates.

The study aimed at demonstrating the performance of a procedure to generate precise LEO orbits within a global GPS data analysis. The orbit model used for JASON is not yet good enough to take full profit of the inclusion of the LEO for the estimated global parameters. A potential for improvements may be viewed in the fact that the effects observed in the global parameters – in particular for GPS orbits and geocenter coordinates – are larger than expected from adding a single receiver to the 120 station ground network.

The quality of the reconstructed JASON orbit does currently not justify to start including LEOs into the global processing at the CODE Analysis Center. Further studies based on longer data spans will identify the origin of the observed problems and quantify the magnitude of improvements that could be expected by performing a combined processing of GPS and LEO orbits together with all global parameters.

References

Beutler, G., E. Brockmann, W. Gurtner, U. Hugentobler, L. Mervart, M. Rothacher (1994). Extended orbit modeling techniques at the CODE processing center of the international GPS service for geodynamics (IGS): theory and initial results. *Manuscripta Geodetica*, 19, pp. 367-386.

Boomkamp, H. (2002). The CHAMP orbit comparison campaign. In: *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*, C. Reigber et al. (eds), pp. 53-58, Springer, Berlin, ISBN 3-540-00206-5. http://nng.esoc.esa.de/gps/igsleo.html

Ferland, R. (2002). Activities of the international GPS Service (IGS) Reference Frame Working Group. in *Vistas for Geodesy in the New Millennium*, J. Adam et al. (eds), International Association of Geodesy Symposia 125, pp. 3-8, Springer, 2002.

Haines, B., W. Bertiger, S. Desai, D. Kuang, T. Munson, L. Young, P. Willis (2002). Initial Orbit Determination Results for Jason-1: Towards a 1-cm Orbit. In: *Proceedings of ION GPS 2002*, Portland, OR, September 24-27, 2002. To be published in *Navigation*.

Hugentobler, U., S. Schaer, P. Fridez (eds) (2001). Bernese GPS Software Version 4.2. Astronomical Institute, University of Berne, February 2001.

Hugentobler, U., S. Schaer, T. Springer, G. Beutler, H. Bock, R. Dach, D. Ineichen, L. Mervart, M. Rothacher, U. Wild, A. Wiget, E. Brockmann, G. Weber, H. Habrich, C. Boucher (2003). CODE IGS Analysis Center Technical Report 2002. In: *IGS 2002 Technical Reports*, K. Gowey et al. (eds). IGS Central Bureau, JPL, CA, USA, in preparation.

Hugentobler, U., G. Beutler (2003). Strategies for precise orbit determination for low Earth orbiters using the GPS. In *Earth Gravity Field from Space – From Sensors to Earth Sciences*, edited by G. Beutler et al. (eds). *Space Science Reviews*, Kluwer, Dortrecht, in press.

Rim, H. J., B. E. Schutz, P. A. M. Abusali, B. D. Tapley (1995). Effects of GPS orbit accuracy on GPS-determined TOPEX/POSEIDON orbit. In: *Proceedings of ION GPS-95*, pp. 613-617, September 12-15, 1995.

Rocken, C., Y.-H. Kuo, W. Wchreiner, D. Hunt, S. Sokolovskiy, C. McCormick (2000). COSMIC System Description. *Terrestrial Atmospheric and Oceanic Sciences*, 11(1), pp. 21-52.

Visser, P. N. A. M., J. van den IJssel (2001). GPS-based precise orbit determination of the very low Earth-orbiting gravity mission GOCE. *J. of Geodesy*, 74, pp. 590-602.

Willis, P., B. Haines, Y. Bar-Sever, W. Bertiger, R. Muellerschoen, D. Kuang, S. Desai, (2003). TOPEX/JASON Combined GPS/DORIS Orbit Determination in the Tandem Phase. *Adv. Space Res.*, 31(8), pp. 1941-1946.

Zhu, S., H. Neumayer, F. H. Massmann, C. Shi, C. Reigber (2002). Impact of different data combinations on the CHAMP orbit determination. In: *Proceedings of the IGS Workshop 2002*, Ottawa, Canada, April 8–12, 2002, in press.

Efficient Stochastic Orbit Modeling Techniques using Least Squares Estimators

A. Jäggi, G. Beutler, U. Hugentobler
Astronomical Institute,
University of Berne, Sidlerstrasse 5, CH-3012 Berne, Switzerland
email: adrian.jaeggi@aiub.unibe.ch

Abstract. Reduced-dynamic orbit determination for spaceborne GPS receivers of low Earth orbiting satellites is a successful method promising highest precision. We review the principles of (reduced) dynamic orbit determination and develop the mathematical background for different efficient stochastic orbit parametrizations (e.g., piecewise constant accelerations which provide not only continuous but also differentiable orbits) using least squares methods. Simulated as well as real data from the CHAMP GPS receiver show, to some extent, the equivalence of the different parametrizations and reveal the impressive performance of stochastic orbit modeling techniques. Independent comparisons with orbits determined by other groups and validations with SLR measurements show that our orbits are of high quality.

Key words. Low Earth orbiter, reduced-dynamic orbit determination, stochastic orbit parametrization

1 Introduction

The idea of using spaceborne GPS receivers for precise orbit determination (POD) of low Earth orbiters (LEOs) was demonstrated for the first time almost a decade ago (Bertiger, 1994) for dynamic and reduced-dynamic orbits of the TOPEX/POSEIDON satellite. This proof of concept stimulated an increasing number of follow-up missions of Earth observing satellites to carry on-board GPS receivers, not only for POD, but also for purposes like atmosphere sounding (Kursinski, 1997) and gravity field recovery. For the latter purpose the analysis of CHAMP GPS receiver and accelerometer data proved that it is possible to separate gravitational from non-gravitational perturbations (Reigber, 2002).
This article focuses on POD, namely on the promising approach of reduced-dynamic orbit determination and in particular on the stochastic orbit modeling part, which defines the degree of "strength" reduction of the dynamic laws by allowing for a stochastic component in the equations of motion. The results for simulated and real GPS data of the CHAMP GPS receiver underline the applicability of the theoretical developments.

2 Dynamic Orbit Determination

The equation of motion of an Earth orbiting satellite including all perturbations reads as

$$\ddot{\boldsymbol{r}} = -GM\frac{\boldsymbol{r}}{r^3} + \boldsymbol{f}_1(t, \boldsymbol{r}, \dot{\boldsymbol{r}}, q_1, ..., q_d) \doteq \boldsymbol{f} \qquad (1)$$

in the inertial frame with initial conditions $\boldsymbol{r}^{(k)}(t_0) = \boldsymbol{r}^{(k)}(a, e, i, \Omega, \omega, T_0; t_0)$, $k = 0, 1$ where $a, e, i, \Omega, \omega, T_0$ are the six osculating elements pertaining to t_0 and where $q_1, ..., q_d$ are the unknown dynamical parameters describing the perturbing acceleration acting on the satellite (e.g., gravity field coefficients, air-drag or radiation pressure parameters, ...).
Let us assume that an a priori orbit $\boldsymbol{r_0}(t)$ is available (e.g., from a GPS code solution). As $\boldsymbol{r_0}(t)$ must be a solution of the equation of motion (1) a priori values $p_{i0} \in \{a_0, e_0, i_0, \Omega_0, \omega_0, T_{00}, q_{10}, ..., q_{d0}\}$ for the unknown orbit parameters p_i are available, as well. The actual orbit $\boldsymbol{r}(t)$ may therefore be linearized by a truncated Taylor series

$$\boldsymbol{r}(t) = \boldsymbol{r_0}(t) + \sum_{i=1}^{n} \frac{\partial \boldsymbol{r_0}(t)}{\partial p_i} \cdot (p_i - p_{i0}) \,, \qquad (2)$$

where $n = 6+d$ denotes the total number of unknown orbit parameters $\{p_1, ..., p_{6+d}\} = \{a, e, ..., q_d\}$. Equation (2) allows it to improve the orbit if the partial derivatives of the a priori orbit w.r.t. the unknown orbit parameters are known.

2.1 Variational Equations

Let us assume that p is one of the parameters defining the initial values or the dynamics in Eq. (1) and that the partial derivative of the orbit $\boldsymbol{r_0}(t)$ w.r.t. the parameter p is designated by the function

$$\boldsymbol{z_p}(t) \doteq \frac{\partial \boldsymbol{r_0}(t)}{\partial p} \,. \qquad (3)$$

The initial value problem associated with the partial derivative (3) is referred to as the system of variational equations in this article and may be obtained

by taking the partial derivative of Eq. (1). The result may be written as

$$\ddot{\boldsymbol{z}}_{\boldsymbol{p}} = \boldsymbol{A_0} \cdot \boldsymbol{z_p} + \boldsymbol{A_1} \cdot \dot{\boldsymbol{z}}_{\boldsymbol{p}} + \frac{\partial \boldsymbol{f_1}}{\partial p} , \tag{4}$$

where the 3×3 matrices $\boldsymbol{A_0}$ and $\boldsymbol{A_1}$ are defined by

$$A_{0[ik]} = \frac{\partial f_i}{\partial r_{0,k}}, \quad A_{1[ik]} = \frac{\partial f_i}{\partial \dot{r}_{0,k}} , \tag{5}$$

where f_i denotes the i-th component of the total acceleration $\boldsymbol{f}$ in Eq. (1). For $p \in \{a, e, i, \Omega, \omega, T_0\}$ Eq. (4) is a linear, homogeneous, second order differential equation system with initial values $\boldsymbol{z_p}(t_0) \neq \boldsymbol{0}$ and $\dot{\boldsymbol{z}}_{\boldsymbol{p}}(t_0) \neq \boldsymbol{0}$, whereas for $p \in \{q_1, ..., q_d\}$ Eq. (4) is inhomogeneous with zero initial values. Note that in the latter case the homogeneous part of Eq. (4) is the same as for parameters p defining the initial values.

3 Pseudo-Stochastic Orbit Modeling

Apart from the six initial conditions, accelerations are set up as the only orbit parameters. Without taking into account any a priori models for the non-conservative forces, piecewise constant stochastic accelerations partially absorb these accelerations (e.g., air-drag, radiation pressure, ...) as well as accelerations due to a mismodeling in the gravitational attractions (e.g., due to an imperfect gravity field model). The parametrization subsequently used for the residual acceleration $\boldsymbol{a}(t)$ in the inertial system reads as the sum of three accelerations in (r)adial, (a)long-track and (c)ross-track directions in the satellite co-rotating system

$$\boldsymbol{a}(t) = \boldsymbol{a_r}(t) + \boldsymbol{a_a}(t) + \boldsymbol{a_c}(t) , \tag{6}$$

where the acceleration $\boldsymbol{a_d}(t)$, $d \in \{r, a, c\}$ is represented by

$$\boldsymbol{a_d}(t) = (a_{0,d} + \sum_{i=1}^{m} a_{i,d}\xi(t, t_{i-1}, t_i)) \cdot \boldsymbol{e_d}(t) , \tag{7}$$

where $\boldsymbol{e_d}$ denotes the unit vector in one of the directions specified above and

$$\xi(t, t_{i-1}, t_i) = \begin{cases} 1 \; ; \; t_{i-1} \leq t < t_i \\ 0 \; ; \; \text{else} \end{cases} . \tag{8}$$

Altogether a total number of $6+3(m+1)$ orbit parameters are set up. As opposed to the constant acceleration $a_{0,d}$ acting over the whole arc in Eq. (7), the m piecewise constant accelerations $a_{i,d}$ are referred to as pseudo-stochastic accelerations in this article, because they are characterized by an expectation value of zero and an a priori weight $w_{i,d}$ given by an a priori variance $\sigma^2_{a_{i,d}}$

$$E(a_{i,d}) = 0 \quad ; \quad w_{i,d} = \frac{\sigma_0^2}{\sigma^2_{a_{i,d}}} , \tag{9}$$

where σ_0 denotes the a priori root mean square error (RMS) of unit weight. The a priori weights in Eq. (9) constrain the estimates of the piecewise constant accelerations $a_{i,d}$, not allowing them to deviate too much from zero. Note that Eq. (9) allows for an estimation of the parameter $a_{0,d}$ in Eq. (7).
As indicated in section 2.1, every orbit parameter additionally set up requires in principle the simultaneous integration of one additional differential equation system (4). Facing the possibly large number of stochastic accelerations to be set up, it is mandatory to develop efficient methods to calculate the partial derivatives w.r.t. these parameters in order to avoid inefficiency due to a computational overload.

3.1 Pseudo-Stochastic Accelerations

Let us now develop the mathematical background for estimating m constant accelerations $a_{i,d}$ in predetermined directions $\boldsymbol{e_d}(t)$ for $t_{i-1} \leq t < t_i, i = 1, ..., m$. For the sake of simplicity we drop the indices i, d in the following and focus on one acceleration a. The contribution of this parameter $q_i = a$ in $\boldsymbol{f_1}$ in Eq. (1) is of the form $a \cdot \boldsymbol{e}(t)$ for $t_{i-1} \leq t < t_i$. Neglecting any possible velocity-dependent forces, the corresponding variational equation thus reads as

$$\ddot{\boldsymbol{z}}_{\boldsymbol{a}} = \boldsymbol{A_0} \cdot \boldsymbol{z_a} + \begin{cases} \boldsymbol{e}(t) \; ; \; t_{i-1} \leq t < t_i \\ \boldsymbol{0} \quad ; \; \text{else} \end{cases} . \tag{10}$$

Let us assume that the functions $\boldsymbol{z_j}(t), j = 1, ..., 6$ are the partial derivatives of the a priori orbit w.r.t. the six parameters defining the six initial conditions at t_0. As these six functions $\boldsymbol{z_j}(t)$ form one complete system of homogeneous solutions of Eq. (10) the solution of the inhomogeneous system (10) and its first time derivative can be obtained by the method of variation of constants and may therefore be written as a linear combination of the homogeneous solutions

$$\boldsymbol{z_a}^{(k)}(t) = \sum_{j=1}^{6} \alpha_j(t) \boldsymbol{z_j}^{(k)}(t) \; ; \; k = 0, 1 , \tag{11}$$

where the coefficients $\alpha_j(t)$ are functions of time t to be determined. Introducing the matrix notation $\boldsymbol{Z_{[1,...,3;j]}} = \boldsymbol{z_j}, \boldsymbol{Z_{[4,...,6;j]}} = \dot{\boldsymbol{z}}_{\boldsymbol{j}}, j = 1, ..., 6$ and vector notation $\boldsymbol{h^T} = (\boldsymbol{0^T}, \boldsymbol{e}(t)^T)$ the solution may be obtained by definite integrals

$$\boldsymbol{\alpha}(t) = \int_{t_0}^{t} \boldsymbol{Z}^{-1}(t')\boldsymbol{h}(t')dt' = \int_{t_{i-1}}^{t^*} \boldsymbol{Z}^{-1}(t')\boldsymbol{h}(t')dt' , \tag{12}$$

where $\boldsymbol{\alpha}^T = (\alpha_1, ..., \alpha_6)$ and where

$$t^* = \begin{cases} t_{i-1} & ; t < t_{i-1} \\ t & ; t_{i-1} \le t < t_i \\ t_i & ; t \ge t_i \end{cases} . \quad (13)$$

This implies that the solution $\boldsymbol{z_a}(t)$ for the parameter a reads as

$$\boldsymbol{z_a}(t) = \begin{cases} \mathbf{0} & ; t < t_{i-1} \\ \sum_{j=1}^{6} \alpha_j(t)\boldsymbol{z_j}(t) & ; t_{i-1} \le t < t_i \\ \sum_{j=1}^{6} \alpha_j(t_i)\boldsymbol{z_j}(t) & ; t \ge t_i \end{cases} . \quad (14)$$

Note that in the case $t \ge t_i$ the coefficients $\alpha_j(t_i)$ are constant in time.

3.1.1 Efficient Solution for $t_{i-1} \le t < t_i$

Let us introduce an auxiliary problem and write the function (14) as a function of the solution of this auxiliary problem. The parameter underlying the auxiliary problem is a constant acceleration $\bar{a}$ over the entire arc. Note that $\bar{a}$ can be identified with $a_{0,d}$ in Eq. (7). The corresponding variational equation (see Eq. (10)) reads as

$$\ddot{\boldsymbol{z}}_{\bar{\boldsymbol{a}}} = \boldsymbol{A_0} \cdot \boldsymbol{z}_{\bar{\boldsymbol{a}}} + \boldsymbol{e}(t) \quad \forall t . \quad (15)$$

As the difference $\boldsymbol{\Delta} = \boldsymbol{z}_{\bar{\boldsymbol{a}}} - \boldsymbol{z_a}$ solves the homogeneous differential equation system $\ddot{\boldsymbol{\Delta}} = \boldsymbol{A_0} \cdot \boldsymbol{\Delta}$ in the designated time interval, its solution can be written as a linear combination of the functions $\boldsymbol{z_j}(t)$, therefore

$$\boldsymbol{z_a}^{(k)}(t) = \boldsymbol{z}_{\bar{\boldsymbol{a}}}^{(k)}(t) - \sum_{j=1}^{6} \beta_j \boldsymbol{z_j}^{(k)}(t) \; ; \; k = 0, 1 . \quad (16)$$

Evaluating Eq. (16) at time t_{i-1} and taking into account Eq. (14), the coefficients β_j may be obtained as a solution of the following linear system of algebraic equations:

$$\begin{aligned} \sum_{j=1}^{6} \beta_j \boldsymbol{z_j}(t_{i-1}) &= \boldsymbol{z}_{\bar{\boldsymbol{a}}}(t_{i-1}) \\ \sum_{j=1}^{6} \beta_j \dot{\boldsymbol{z}}_{\boldsymbol{j}}(t_{i-1}) &= \dot{\boldsymbol{z}}_{\bar{\boldsymbol{a}}}(t_{i-1}) . \end{aligned} \quad (17)$$

As the above equations form a linear system of six scalar equations for the six unknowns β_j it is possible to write the partial derivative (16) w.r.t. the parameter a as a linear combination of the partial derivative w.r.t. the parameter $\bar{a}$ and the six partial derivatives $\boldsymbol{z_j}(t)$. Note that Eq. (11) together with Eq. (12) implies not only continuity but also differentiability of $\boldsymbol{z_a}(t)$ at time t_{i-1}.

3.1.2 Efficient Solution for $t \ge t_i$

Equations (11) and (14) imply that $\boldsymbol{z_a}(t)$ is continuous and differentiable at time t_i, as well, which means that $\boldsymbol{z_a}^{(k)}(t_i)$ may be computed by evaluating Eq. (16) at time t_i. The coefficients $\alpha_j(t_i)$ may therefore be obtained as a solution of the following linear system of algebraic equations:

$$\begin{aligned} \sum_{j=1}^{6} \alpha_j(t_i)\boldsymbol{z_j}(t_i) &= \boldsymbol{z_a}(t_i) \\ \sum_{j=1}^{6} \alpha_j(t_i)\dot{\boldsymbol{z}}_{\boldsymbol{j}}(t_i) &= \dot{\boldsymbol{z}}_{\boldsymbol{a}}(t_i) . \end{aligned} \quad (18)$$

As the above equations again form a linear system of six scalar equations for the six unknowns $\alpha_j(t_i)$ it is possible to write the partial derivative $\boldsymbol{z_a}(t)$ as a linear combination of the six partial derivatives $\boldsymbol{z_j}(t)$ only. In conclusion not only the position vector $\boldsymbol{r}(t)$ but also the velocity vector $\dot{\boldsymbol{r}}(t)$ of the improved orbit remains continuous over the whole arc.

3.2 Pseudo-Stochastic Pulses

Let us briefly outline the special case of instantaneous velocity changes v_i at times t_i in predetermined directions $\boldsymbol{e}(t_i)$ (Beutler, 1994). Focusing on one pulse v at time t_i, the contribution of $q_i = v$ in $\boldsymbol{f_1}$ in Eq. (1) may formally be written as $v \cdot \delta(t - t_i) \cdot \boldsymbol{e}(t)$ where $\delta(t)$ denotes Dirac's delta function. The corresponding variational equation thus reads as

$$\ddot{\boldsymbol{z}}_{\boldsymbol{v}} = \boldsymbol{A_0} \cdot \boldsymbol{z_v} + \delta(t - t_i)\boldsymbol{e}(t) . \quad (19)$$

Using the same notation the solution of (19) may be written again in the form of Eq. (11). The coefficients $\alpha_j(t)$ may be obtained in analogy to Eq. (12) by definite integrals which are given for $t \ge t_i$ by

$$\boldsymbol{\alpha}(t) = \int_{t_0}^{t} \delta(t' - t_i)\boldsymbol{Z}^{-1}(t')\boldsymbol{h}(t')dt' = \boldsymbol{Z}^{-1}(t_i)\boldsymbol{h}(t_i) \quad (20)$$

Obviously the $\alpha_j(t)$ are constant in time for $t \ge t_i$. Therefore the partial derivatives $\boldsymbol{z_v}(t)$ may be written as a linear combination of the six partial derivatives w.r.t. the initial conditions only. The drawback however is that $\dot{\boldsymbol{z}}_{\boldsymbol{v}}(t)$ and therefore $\dot{\boldsymbol{r}}(t)$ are no longer continuous. The discontinuities lie at the epochs t_i.

4 Simulation Studies

We discuss two experiments using simulated GPS phase zero-difference observations for the CHAMP satellite. In both simulations the same physical and mathematical models were used as in the real data

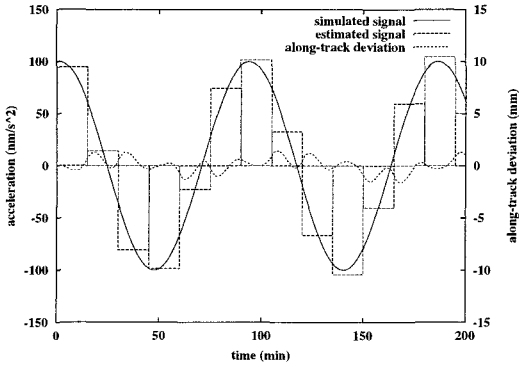


Fig. 1. Piecewise constant accelerations every 15 minutes compensate for an unmodeled along-track signal. The dotted curve denotes deviations w.r.t. the "true" orbit (simulated data).

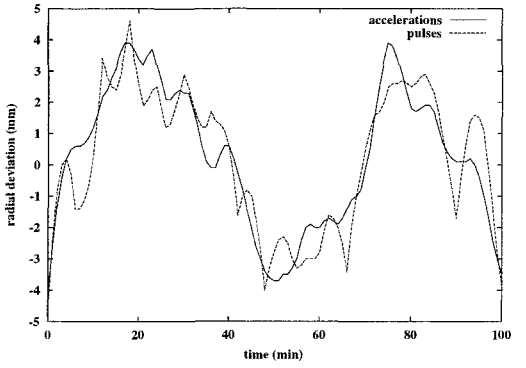


Fig. 2. Radial deviations w.r.t. the "true" orbit for pseudo-stochastic accelerations resp. pulses set up every 6 minutes. Epochs of velocity changes are easily recognized on the dashed curve (simulated data).

processing. In addition the error-free GPS data were affected by an artificial once per rev. along-track acceleration with an amplitude of 10^{-7} m/s^2 for the first simulation. An orbit with identical initial values without being affected by the artificial signal served as a priori orbit. To illustrate the feasibility of our approach for this scenario, piecewise constant accelerations in along-track direction were estimated every fifteen minutes. For a selected time interval of about two revolution periods, Fig. 1 shows that the estimated piecewise constant accelerations follow the artificial ("true") signal and that the deviations from the "true" orbit in along-track direction are mostly below the one millimeter level. High correlations with the time intervals of the accelerations can be recognized as well as a once per rev. periodicity. Both effects are caused by the unavoidable deficiencies of the parametrization.

The second simulation illustrates the applicability and, to some extent, the equivalence of both, pseudo-stochastic accelerations and pulses in a more realistic environment. Again GPS observations were simulated but as opposed to the first experiment the phase observations were affected by a white noise random error of 1 mm RMS. A dynamical orbit using the gravity field model EIGEN-1S (Reigber, 2002) up to degree/order 120 served as "true" orbit, whereas the same gravity field model, but truncated at degree/order 80, was used in the orbit improvement procedure. To illustrate the properties of both parametrizations two different orbits were generated, one with the parametrization defined in Eq. (6) (where stochastic accelerations were set up every six minutes) and another with pulses instead of accelerations with the same spacing in time. By optimizing the constraints in Eq. (9), both orbits could be fitted equally well (postfit RMS of about 1.02 mm). The differences w.r.t. the "true" orbit yield almost the same RMS values over one day, smaller than 2 mm in all three directions. The largest differences occur in the radial component (caused by the large number of rather loosely constrained stochastic parameters, which reduce the dynamic to the extent that the resulting orbit tends to be more sensitive to the observation geometry (like a kinematic orbit) which penalizes the radial direction). Fig. 2 shows the radial deviations w.r.t. the "true" orbit over a selected time interval of about one revolution period where the differences for both parametrizations can be observed well. The orbit generated with piecewise constant accelerations shows a smooth behaviour, whereas epochs with velocity changes can be identified quite easily on the other orbit. Nevertheless, the differences are very small (below a few millimeters in the vicinity of the epochs of the stochastic parameters) and do not significantly affect the overall orbit quality. Consequently, no significant gain or loss has to be expected using one or the other orbit parametrization with a reasonably large number of stochastic parameters, when processing real data. The real errors are much higher due to the non-gravitational forces, which were neglected in this simulation, and therefore exceed the subtleties reported above.

5 CHAMP Orbit Test Campaign

Using GPS final orbits and high-rate clocks (Bock, 2002) from the CODE analysis center (Center for Orbit Determination in Europe) real CHAMP GPS zero-difference phase observations were processed to study the impact of constrained stochastic parameters on the orbit. The orbit quality was assessed by internal (e.g., formal accuracies along the orbit) and external indicators (e.g., comparisons with orbits from

Table 1. RMS of plain differences to an orbit from CSR resp. RMS of SLR residuals (day 141).

Constraint (m/s^2)	RMS w.r.t. CSR (cm)	RMS of SLR residuals (cm)
$5 \cdot 10^{-8}$	6.72	3.07
$1 \cdot 10^{-8}$	5.09	1.77
$5 \cdot 10^{-9}$	4.89	1.99
$1 \cdot 10^{-9}$	5.32	3.38

other groups, SLR residuals). For this purpose we used the time span of the CHAMP orbit test campaign from May 20 to 30, 2001 (days 140-150), which was selected as the test period within the IGS LEO Pilot Project. As the orbits from the Center of Space Research (CSR) at Austin were found to be the best of the twelve contributing centers (Boomkamp, 2002) they serve here as benchmark orbits.

5.1 Internal Quality Assessment

In order to assess the orbit quality, the full parametrization (defined in Eq. (6)) with stochastic parameters set up every six minutes was used to generate CHAMP orbits for both, accelerations and pulses with different (but in each direction equal) constraints. To obtain best results, the gravity field model EIGEN-1S was used and the attitude measurements from the star sensors were introduced. The dependency of the postfit RMS on the constraints was found to be equal for both parametrizations, obviously allowing a better fit for looser constraints. However this does not necessarily imply that such orbits are really better. On the contrary, it turns out that setting up a large number of loosely constrained stochastic parameters may result in worse orbits, which may be verified by an external comparison. Table 1 shows for day 141 the RMS of plain 1-d differences (no Helmert transformation applied) to an orbit from CSR as well as the RMS of SLR residuals for four different types of constraints (in each direction equal) when setting up stochastic accelerations every six minutes. According to this comparison an optimal solution exists in the range considered and most probably the constraints of $5 \cdot 10^{-9}$ m/s^2 are quite close to the optimum. Note that almost identical results can be obtained if pulses are set up.
It is instructive to analyze the formal errors along the orbit as an additional indicator of quality. Applying the general law of error propagation together with the variance-covariance information allows the computation of formal errors in the system co-rotating with the satellite (i.e., in radial, along-track and cross-track direction). Fig. 3 shows as an example the errors in radial direction for the four different constraints from Table 1. Apart from the common characteristics (e.g., the increased error level around 3h and 10h caused by

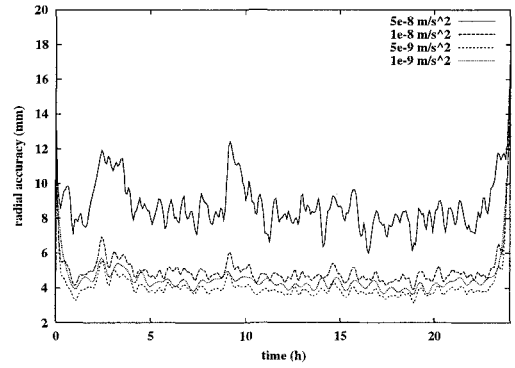


Fig. 3. Formal accuracies of orbit positions in radial direction for differently constrained pseudo-stochastic accelerations set up every 6 minutes using the gravity field model EIGEN-1S (day 141).

a reduced number of successfully tracked satellites), Fig. 3 implies that good solutions in Table 1 turn out to have low formal accuracies along the orbit as well. The accuracies in the other directions show a similar behaviour, whereas the highest error level can be found in the along-track direction indicating that the orbits (as opposed to the simulations in section 4) are all governed by the dynamical laws due to the relatively tight constraints. Nevertheless the "most" kinematic solution with constraints of $5 \cdot 10^{-8}$ m/s^2 already suffers considerably from the limit of the CHAMP BlackJack receiver at that time to track only up to 8 satellites simultaneously.

5.2 External Quality Assessment

Table 2 shows the daily RMS of plain 1-d differences w.r.t. CSR orbits (middle column) for the time span of the CHAMP orbit comparison test campaign. Over the eleven days an RMS of 5.29 cm w.r.t. the orbits from CSR (corresponding to a 3-d RMS of 9.16 cm) is achieved. Solving for a seven parameter Helmert transformation shows that our orbits agree with the CSR orbits on a RMS level better than 5 cm. However both orbit types show a mean shift in the earth-fixed z-direction of 3.32 cm which may be due to slightly different reference frames the used GPS orbits and clocks refer to.
Table 2 shows the daily RMS of SLR residuals (right column) for the time span of the CHAMP orbit comparison test campaign. A total of 1837 SLR measurements (normal points) of 16 stations were used for this validation. Due to the sparse SLR tracking the differences of SLR measurements (corrected for the tropospheric delay) and our orbit positions give only snapshots of the 1-d orbit accuracy in selected directions during the short tracking passes. For the eleven

Table 2. RMS of plain differences to orbits from CSR resp. RMS of SLR residuals (days 140-150). An overall RMS of 5.29 cm resp. 3.37 cm results.

Day of year	RMS w.r.t. CSR (cm)	RMS of SLR residuals (cm)
140	4.87	3.19
141	4.89	1.99
142	5.85	2.77
143	5.19	2.51
144	4.91	4.99
145	6.07	4.61
146	4.85	4.15
147	5.12	2.47
148	4.73	2.81
149	5.71	3.08
150	5.74	3.30

days an RMS of 3.37 cm over all SLR residuals is obtained without showing any significant SLR bias.

5.3 Comparison with Accelerometer Data

If a perfect gravity field could be used the estimated accelerations should directly represent the non-conservative forces acting on the satellite. Fig. 4 shows for a selected time interval of about three revolution periods how the piecewise constant accelerations in along-track direction agree with the measured accelerations (bias and scale applied) from the STAR accelerometer, when the gravity field model EIGEN-1S is used. Despite the crude approximation with constant accelerations over six minutes the correlation (77.8%) is quite remarkable whereas the agreement in the other directions is less significant (correlation of 36.1% resp. 34.7% in radial resp. cross-track direction). At least in along-track direction reliable bias and drift parameters may be derived by such comparisons (-3.518 μms^{-2} and 0.0013 μms^{-2}/day for the time span of the orbit comparison campaign) for internal validation purposes.

6 Summary

We developed the mathematical background for a stochastic orbit modeling in the environment of least squares estimators and presented two possible pseudo-stochastic orbit parametrizations (piecewise constant accelerations resp. instantaneous velocity changes) which may be set up in an efficient way. Due to the density and high accuracy of GPS phase tracking data both parametrizations are well suited for reduced-dynamic LEO POD and may even be considered to some extent as equivalent as shown with simulations. The processing of CHAMP zero-difference phase data (days 140-150 in 2001) showed that the estimated reduced-dynamic orbits may be validated with SLR residuals on a level of 3.37 cm RMS without any significant SLR bias and that they agree on

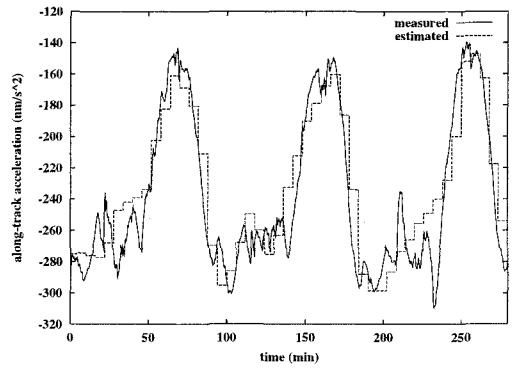


Fig. 4. Piecewise constant accelerations every six minutes in along-track direction compared with measurements from the STAR accelerometer (bias and scale applied) using the gravity field model EIGEN-1S (day 141).

an RMS level of better than 5 cm with orbits from CSR. The additional comparison with accelerometer data showed a high correlation (77.8%) between the measured and the estimated piecewise constant accelerations in along-track direction.

References

Bertiger, W., Y. Bar-Sever, E. Christensen, E. Davis, J. Guinn, B. Haines, R. Ibanez-Meier, J. Jee, S. Lichten, W. Melbourne, R. Muellerschoen, T. Munson, Y. Vigue, S. Wu, T. Yunck, B. Schutz, P. Abusali, H. Rim, M. Watkins, P. Willis (1994). GPS Precise Tracking of TOPEX/POSEIDON: Results and Implication. *Journal of Geophysical Research*, 99(C12): 24 449-24 464.

Beutler, G., E. Brockmann, W. Gurtner, U. Hugentobler, L. Mervart, M. Rothacher (1994). Extended orbit modeling techniques at the CODE processing center of the international GPS service for geodynamics (IGS): theory and initial results. *Manuscripta Geodetica*, 19:367-386

Bock, H., U. Hugentobler, T.A. Springer, G. Beutler (2002). Efficient Precise Orbit Determination of LEO Satellites using GPS. *Advanced Space Research*, 30/2, pp. 295-300.

Boomkamp, H. (2002). The CHAMP Orbit Comparison Campaign. In: *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*, edited by C. Reigber et al., pp. 53-58, Springer, Berlin, ISBN 3-540-00206-5.

Kursinski, E.R., G.A. Hajj, J.T. Schofield, R.P. Linfield, K.R. Hardy (1997). Observing Earth's Atmosphere with Radio Occultation Measurements Using the Global Positioning System. *Journal of Geophysical Research*, 102(D19): 23 429-23 465.

Reigber, Ch., G. Balmino, P. Schwintzer, R. Biancale, A. Bode, J.-M. Lemoine, R. Koenig, S. Loyer, H. Neumayer, J.-C. Marty, F. Barthelmes, F. Perosanz, SY. Zhu (2002). A High Quality Global Gravity Field Model from CHAMP GPS Tracking Data and Accelerometry (EIGEN-1S). *Geophysical Research Letters*, 29(14),10.1029/2002GL015064.

Kinematic Precise Orbit Determination for Gravity Field Determination

D. Švehla and M. Rothacher
Forschungseinrichtung Satellitengeodäsie, Institut für Astronomische und Physikalische Geodäsie,
Technical University of Munich, Arcisstr. 21, D-80290 Munich, Germany. E-mail: svehla@bv.tum.de

Abstract.

In this paper we first present approaches and results in precise orbit determination (POD) for satellites in Low Earth Orbit (LEO) based on one or two frequency GPS measurements and, secondly, we focus on the relations between kinematic POD and gravity field determination. Using GPS measurements of the CHAMP satellite we show that it is possible to estimate kinematic positions of a LEO satellite with the same level of accuracy ($\approx$ 1–3 cm w.r.t. SLR) as with the widely applied reduced-dynamic or dynamic approaches. Kinematic precise orbit determination (POD) as presented here is based on GPS phase measurements and is independent of satellite dynamics (e.g. gravity field, air-drag, etc.) and orbit characteristics (e.g. orbit height, eccentricity, etc.). We also looked at the LEO POD based on GPS measurements from only one frequency, where we make use of what we call the LP linear combination. We show that with this linear combination LEO POD can be performed with one frequency at the 10 cm level. In the second part of the paper the use of kinematic POD is discussed in the framework of the CHAMP, GRACE and GOCE gravity missions. With simulated GPS measurements we studied the impact of ambiguity resolution for the kinematic baseline between the two GRACE satellites. At the end of the paper we present an alternative approach in gravity field determination by measuring the gravitational frequency shift between (optical) atomic clocks in space. In this approach we require very accurate clocks positions, which may be obtained from kinematic POD.

Keywords. Kinematic orbit, LEO, POD, space clock

1 Introduction

With the pioneering satellite mission CHAMP, a new era in space geodesy and observing the planet Earth from space started. Today we are talking about gravity (CHAMP, GRACE, GOCE) and magnetic (CHAMP, ÖRSTED, SWARM) fields determined from space, atmosphere sounding from space (CHAMP, GRACE, COSMIC, SWARM), monitoring oceans (TOPEX/POSEIDON, JASON-1) and ice caps (ICESAT) from space, etc.

In all these missions, satellite orbit determination is used for geo-location of the satellite sensors on one hand and to measure the gravity field and its variations in time on the other hand, i.e., using the equation of motion to obtain information about dynamical processes in the Earth system, as e.g. Earth tides, mass distribution, ocean circulations, etc. (Balmino et al. (1999)).

In 1992, for the first time, high-precision LEO dynamic orbit determination was performed making use of GPS measurements from TOPEX/POSEIDON (Bertiger et al. (1994), Tapley et al. (1994), Yunck et al. (1994)). Since this time, GPS tracking has become an extremely successful method for POD and nowadays purely kinematic orbits can be determined with the same level of accuracy as orbits computed with the more common (reduced-)dynamic approach (Švehla & Rothacher (2002a)). Among all space geodetic techniques (SLR, DORIS, altimetry, etc.) only GPS allows purely kinematic precise orbit determination, where kinematic satellite positions can be estimated independently of orbit altitude and force modeling like, e.g., gravity field, air-drag, solar radiation, etc.

With highly accurate kinematic satellite positions available as pseudo-observations, interesting new methods are being developed nowadays to e.g. determine gravity field parameters, validate dynamical models or derive atmosphere density information.

2 Kinematic and reduced-dynamic POD

2.1 Description of methods

Using GPS measurements the orbit of a LEO satellite can be computed using kinematic as well as (reduced)-dynamic approaches.

The kinematic approach is a purely geometrical approach without using any information on satel-

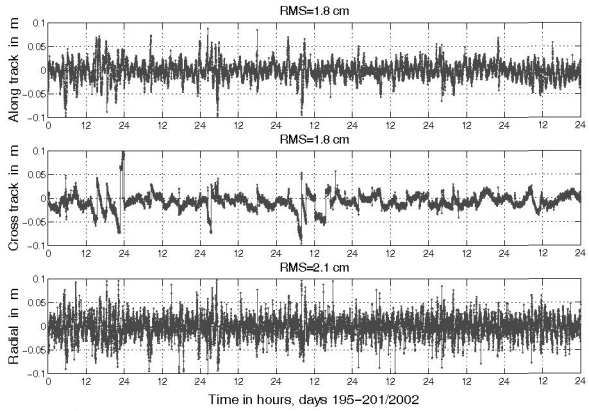


Fig. 1. Differences between CHAMP kinematic and reduced-dynamic orbit based on zero-differences, GPS week 1175/2002 (days 195-201/2002).

lite dynamics (e.g. gravity field, air-drag, etc.). In our case the LEO kinematic orbit is represented by three kinematic coordinates each epoch and estimated with a least-squares adjustment using phase measurements only. Code observations are only used in a pre-processing step to synchronize the LEO clock to GPS time. No constraints are applied to the kinematic positions in the adjustment.

Dynamic and reduced-dynamic POD on the other hand is based on the numerical integration of the equation of motion and the variational equations to obtain the orbit itself, as well as the partial derivatives with respect to the orbital parameters. In the reduced-dynamic case a large number of empirical (e.g. pseudo-stochastic pulses) or force field parameters are estimated in order to cope with the deficiencies in the dynamical models. In our case pseudo-stochastic pulses are set up as additional parameters in along-track, cross-track and radial direction every 6-15 min.

Kinematic or reduced-dynamic POD of a LEO may be carried out on the zero-, the double- or even the triple-difference level of GPS phase or code measurements. In this paper we will focus on zero- and double-difference procedures using phase measurements only.

In the zero-difference case, GPS satellite orbits and clocks are kept fixed and epoch-wise LEO GPS receiver clock parameters are estimated together with either epoch-wise kinematic positions in the kinematic case, or (reduced-)dynamic orbital parameters in the (reduced-)dynamic approach. Since phase GPS measurements are used, more than 400 zero-difference ambiguities are additional parameters for a 1-day orbit arc. POD based on zero-differences is a very fast and efficient approach, because only GPS measurements of the LEO are involved and therefore, the processing of a ground IGS network is only required in a preceding, independent step in order to obtain GPS satellite orbits and clocks.

In the double-difference approach, baselines between the LEO and GPS ground stations are formed and all clock parameters are eliminated. In our case baselines are formed between CHAMP and 40 IGS stations and all IGS products like CODE 3-day solutions for GPS orbits, ERPs, troposphere and weekly station coordinates are kept fixed. The main advantage of kinematic and reduced-dynamic POD based on double-differences is the possibility to resolve ambiguities to their integer values using ambiguity resolution strategies and thereby gain in accuracy. Our ambiguity resolution strategy is based on the Melbourne-Wübbena (MW) linear combination (Melbourne (1985)) to resolve wide-lane ambiguities in a first step and on a subsequent, iterative bootstrapping to resolve narrow-lane ambiguities. 10 narrow-lane ambiguities (out of $\approx$ 5000 ambiguities) are resolved in each iteration step, i.e. before the one-day normal equation matrix is re-inverted.

More details about our zero- and double-difference approaches, ambiguity resolution and the mathematical background of kinematic and reduced-dynamic POD can be found in Švehla & Rothacher (2002a), Švehla & Rothacher (2003b) and Švehla & Rothacher (2003c). Other approaches of kinematic POD can be found, e.g., in Colombo et al. (2002), Bock (2003) or Byun (2003).

2.2 CHAMP POD results

Two comparisons were performed to assess the consistency and accuracy of the CHAMP orbits computed: the comparison between kinematic and reduced-dynamic orbits and the comparison of CHAMP SLR measurements to the GPS-derived orbits.

Fig. 1 shows the difference between CHAMP kinematic and reduced-dynamic orbits based on zero-difference phase measurements for GPS week 1175/2002. The consistency between these two orbits is on the level of about 2 cm over the entire week. The differences between CHAMP kinematic and reduced-dynamic orbits computed using double-differences are displayed in Fig. 2 for day 199/2002. A more detailed look at Fig. 2 reveals that CHAMP kinematic positions sometimes exhibit large spikes, due to the small number of GPS satellites tracked and the resulting poor satellite geometry. These points can easily be recognized when looking at the variance-covariance information. Due to the nature

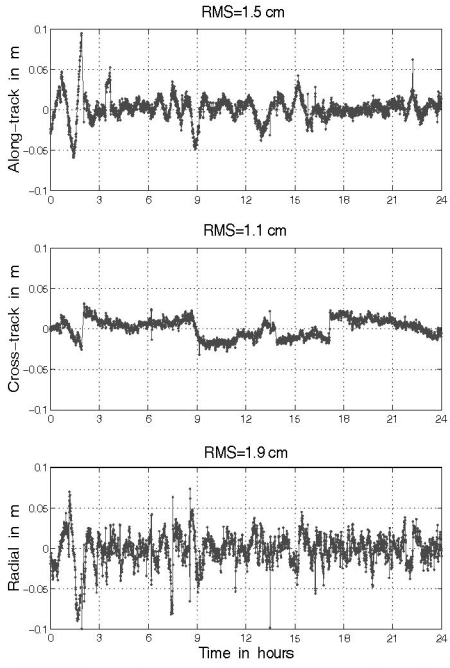


Fig. 2. Difference between CHAMP kinematic and reduced-dynamic orbits based on double-differences, day 199/2002.

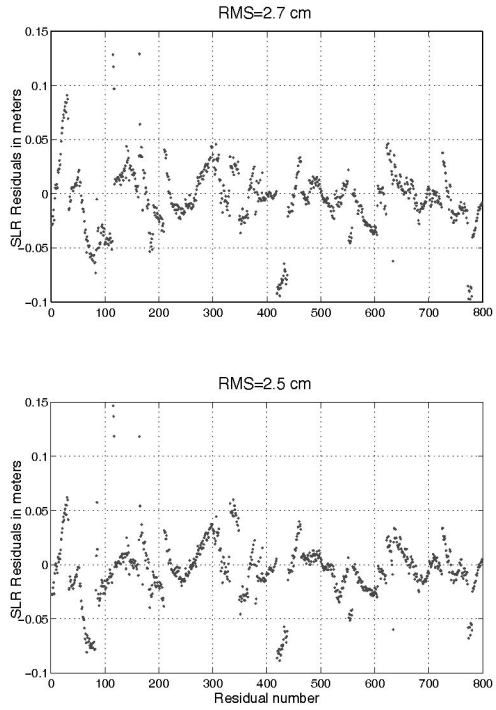


Fig. 3. SLR residuals for CHAMP kinematic (top) and reduced-dynamic orbits (bottom) for GPS week 1175/2002 (days 195-201/2002). All SLR residuals were used in the analysis, elevation cut-off 10°.

of the GPS phase observable, kinematic positions are very smooth from epoch to epoch and as a consequence high-frequency gravity signals may be extracted from these positions. Keep in mind that in our kinematic approach no constraining is applied. Systematic deviations can be recognized in the along-track and radial component pointing, apart from obvious problems with satellite geometry, at deficiencies in the gravity field and air-drag modeling of the reduced-dynamic approach. That we are not just talking about consistency but also accuracy, can be seen in Fig. 3, where SLR residuals are shown for the same kinematic and reduced-dynamic CHAMP orbit as those displayed in Fig. 1. Tropospheric delays for SLR measurements were modeled using the Marini-Murray model and standard corrections like ocean loading (GOT00.2), Shapiro relativistic effect and station velocities were applied. The analysis was performed using ITRF 2000 station coordinates, velocities and eccentricities published by ILRS at http://ilrs.gsfc.nasa.gov/. All SLR stations and SLR measurements were used in this validation (elevation cut-off 10°). Both orbit types exhibit the same quality of about 2.5 cm. It is interesting to note that the SLR residuals show a similar behaviour for kinematic and reduced-dynamic orbits and that no significant bias can be identified in the SLR residuals.

Table 1 summarizes the daily RMS of the SLR residuals for our CHAMP orbits based on four different POD approaches, namely kinematic and reduced-dynamic orbits based on zero- and double-differences. One can see that CHAMP orbits are of similar quality for a purely kinematic and reduced-dynamic approach. This also stands for CHAMP orbits computed using either zero- or double-difference phase measurements. Slightly better orbit quality (2.56 cm) is obtained when using kinematic POD and double-differences.

Day	Zero difference red.-dyn.	Zero difference kinematic	Double difference red.-dyn.	**Double difference kinematic**
195	4.02	4.17	3.22	2.66
196	2.90	2.93	3.19	3.03
197	3.40	3.11	3.29	2.90
198	2.07	2.07	1.99	1.34
199	1.94	1.66	1.91	1.70
200	1.43	1.45	1.69	1.83
201	3.59	4.65	4.32	5.00
202	2.03	2.08	1.93	2.05
Mean	2.67	2.77	2.69	**2.56**

Table 1. Daily RMS of SLR residuals in cm for CHAMP kinematic and reduced-dynamic ("red.-dyn.") orbits based on zero- and double-differences (days 195-202/2002).

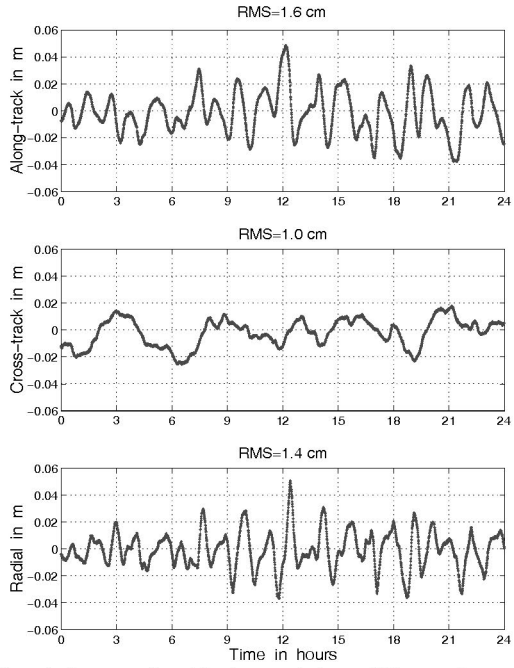


Fig. 4. Impact of ambiguity resolution: difference between reduced-dynamic orbits with float and fixed ambiguities, day 200/2002.

To try to further improve the orbit quality, ambiguity resolution was performed in the double-difference case for GPS week 1175/2002. Using the MW linear combination, about 50% of the wide-lane ambiguities could easily be resolved. These wide-lane ambiguities were introduced in the next step to resolve the narrow-lane ambiguities. Epoch-wise coordinates were pre-eliminated from the normal equation system in kinematic and orbital parameters in reduced-dynamic POD, leaving ambiguities only for bootstrapping. With CHAMP data 20.8% of all ambiguities were resolved in kinematic and 20.4% in reduced-dynamic bootstrapping. Due to the huge number of ambiguity parameters (5000 per day) bootstrapping is very time-consuming and requires about 100 inversions of the 1-day normal equation matrix for both POD approaches. Fig. 4 shows the impact of ambiguity resolution on reduced-dynamic orbits based on double-differences. One can see that ambiguity resolution changes the orbit by 1-2 cm. Unfortunately, the difference between the ambiguity-fixed and ambiguity-free reduced-dynamic orbits is not large enough to give a significantly different results when compared to SLR data.

3 LP linear combination

In this Section the high accuracy of the CHAMP code measurements (used for ambiguity resolution) is assessed by estimating CHAMP orbits using GPS measurements from only one frequency. Following Švehla & Rothacher (2003c), a simplified version of the observation equation for the phase $L^s_{LEO,i}$ and code $P^s_{LEO,i}$ observations (frequency i, between LEO and GPS satellite s) is given as

$$L^s_{LEO,i} = \rho^s_{LEO} + \lambda_i N^s_{LEO,i} - \delta\rho^s_{ion,i} + \\ + c\,\delta t_{LEO} - c\,\delta t^s + \epsilon(L_i) \quad (1)$$
$$P^s_{LEO,i} = \rho^s_{LEO} + \delta\rho^s_{ion,i} + c\,\delta t_{LEO} - c\,\delta t^s + \epsilon(P_i)$$

where ρ^s_{LEO} denotes the distance between LEO and GPS satellite s, $N^s_{LEO,i}$ the zero-difference ambiguity of wavelength λ_i, $\rho^s_{ion,i}$ the ionospheric correction, $c\,\delta t_{LEO}$ and $c\,\delta t^s$ the LEO and GPS satellite clock values and $\epsilon(L_i)$ and $\epsilon(P_i)$ the phase and code noise. The LP linear combination of phase and code measurements on frequency i is then defined as

$$LP^s_{LEO,i} := \frac{1}{2}(L^s_{LEO,i} + P^s_{LEO,i}). \quad (2)$$

See also Bertiger et al. (1996), where this linear combination is called "graphic data". The ionosphere effects in $L^s_{LEO,i}$ and $P^s_{LEO,i}$ cancel and we obtain

$$LP^s_{LEO,i} = \rho^s_{LEO} + \frac{1}{2}\lambda_i N^s_{LEO,i} + \\ + c\,\delta t_{LEO} - c\,\delta t^s + \epsilon(LP_i). \quad (3)$$

We note that the wavelength of the LP linear combination is half the size of the original wavelength λ_i and that the noise $\epsilon(LP_i)$ is a factor of 2 smaller than the code noise

$$\epsilon(LP_i) \approx \frac{1}{2}\epsilon(P_i) \approx 5 - 8\,cm. \quad (4)$$

Since the accuracy of the CHAMP ionosphere-free P-code measurements is about 48 cm (derived from kinematic POD results), we expect the accuracy of the P_1-code measurements to be about 10–16 cm, resulting in a noise level of about 5–8 cm for the CHAMP LP_1 observable. Fig. 5 shows the CHAMP reduced-dynamic orbit we obtain for day 200/2002 using the LP_1 linear combination on the zero-difference level compared to our best reduced-dynamic orbit (double-difference, ionosphere-free phase data). This indicates that in principal LEO orbits with 10 cm accuracy can be obtained using one frequency data only.

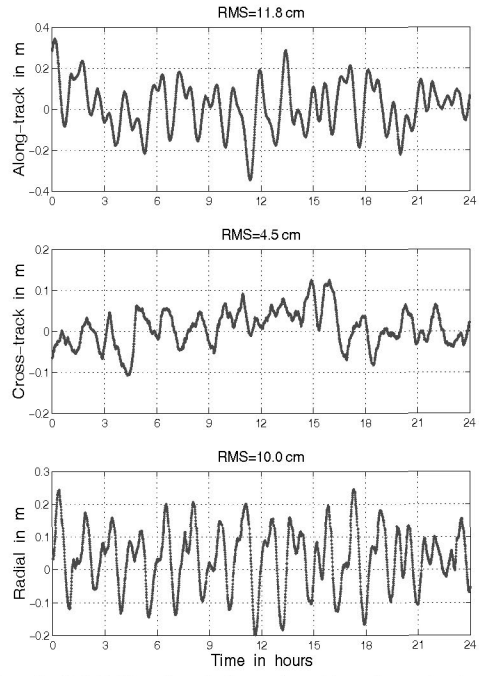


Fig. 5. CHAMP reduced-dynamic orbit estimated using the LP linear combination of the L_1 and P_1 measurements, day 200/2002, showing that LEO orbits can be determined with an accuracy of 10 cm based on single-frequency data only.

The disadvantage of the using LP linear combination and one frequency only is that pre-processing has to be performed with single-frequency data, which might be more difficult. In our analysis preprocessing was done based on dual-frequency data.

4 Gravity field determination based on kinematic orbits

The use of kinematic positions together with their variance-covariance information as an interface to gravity field determination avoids the simultaneous adjustment of gravity field coefficient together with a huge amount of global GPS parameters, like GPS satellite orbits/clocks, zero- or double-difference ambiguities, station coordinates, troposphere parameters, Earth rotation parameters, etc. Comparison with reduced-dynamic orbits and external validation with SLR show that, due to the nature of the phase observable, kinematic positions are very smooth from epoch to epoch (see Fig. 2 and Fig. 3) as long as there are no phase breaks. As a consequence, high-frequency gravity signals may be extracted from these positions. An elegant way to derive gravity field coefficients from kinematic positions is the use of the energy conservation law which may be written in an inertial frame as

$$V = \frac{1}{2}(\frac{d\vec{x}}{dt})^2 - \int_{\vec{x}} \vec{a}_t d\vec{x} - \int_{\vec{x}} \vec{a}_{non} d\vec{x} - C. \quad (5)$$

with the gravitational potential V, the accelerations $\vec{a}_t$ and $\vec{a}_{non}$ due to the time-varying part of the gravity field (e.g. tides) and the non-gravitational forces, respectively, and the total energy constant C. A similar formulation in the earth-fixed frame can be found in Gerlach et al. (2003). An advantage of the gravity field determination based on the energy integral is that we can directly work with the gravity potential as a scalar field instead of having to integrate the equation of motion and all variational equations.

Whereas $\vec{a}_t$ can be obtained from models, the non-gravitational accelerations $\vec{a}_{non}$ are measured by accelerometers. The kinematic energy of the satellite can be computed using velocities derived from kinematic positions by numerical differentiation procedures. At the moment we are providing CHAMP kinematic positions with a sampling of 30 s, which means that the spatial resolution of the estimated gravity field is limited to about 230 km and that much care has to be taken when deriving kinematic velocities. Going to a higher sampling rate, numerical differentiation will become more accurate and a higher spatial resolution will become possible. Making use of the energy integral and kinematic positions from one year of CHAMP data, several gravity field models have recently been computed at the TU Munich, among them the models TUM-1S and TUM-2S (see Gerlach et al. (2003), Földváry et al. (2003)).

In order to get an indication of the quality of these models, CHAMP, JASON-1 and LAGEOS-1 orbits were determined based on GPS (CHAMP, JASON-1) and SLR measurements (LAGEOS-1) using the TUM-1S gravity model. For CHAMP and JASON-1 Table 2 shows the mean RMS of the SLR validation (days 196-202/2002), whereas for LAGEOS-1 the a posteriori RMS of the SLR observations when

Model	JASON-1	CHAMP	LAGEOS-1
EIGEN-1S	4.89	2.24	1.61
EIGEN2	4.69	2.26	1.59
TUM-1S	**4.73**	**2.46**	**1.79**
JGM3	4.28	3.26	1.62
GRIM5-C1	4.96	3.35	1.60
GRIM5-S1	5.03	4.87	1.61

Table 2. Second and third column: Mean RMS of SLR residuals (days 196-202/2002) in cm for CHAMP and JASON-1 GPS-derived orbits based on different gravity models. Last column: A posteriori RMS for LAGEOS-1 dynamic orbits, 8-day arc (days 010-017/1999).

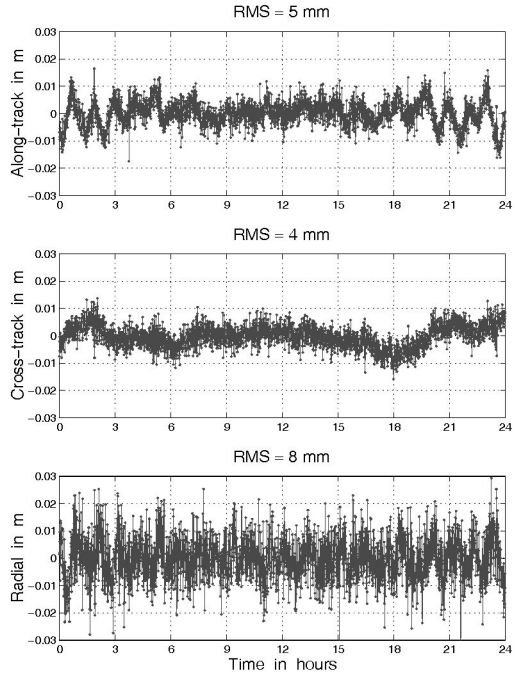


Fig. 6. Kinematic positions of GRACE-A w.r.t. GRACE-B from simulated data with float ambiguities compared to true baseline. Notice colored noise, reflecting correlations between positions and ambiguities.

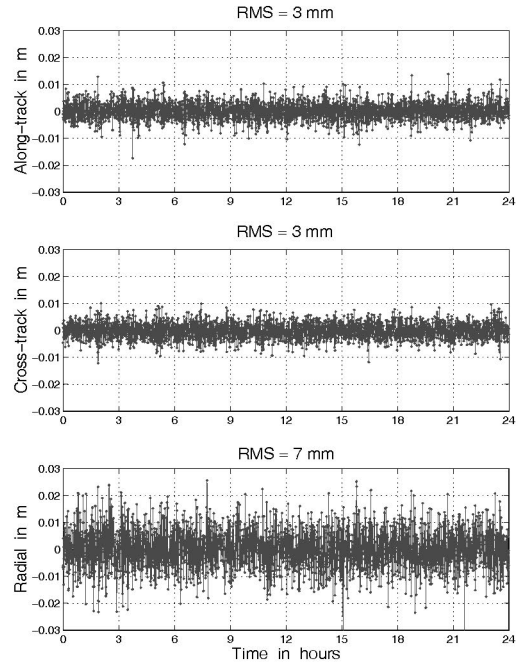


Fig. 7. Kinematic positions of GRACE-A w.r.t. GRACE-B from simulated data with resolved ambiguities compared to true baseline. Notice white noise in the kinematic positions and reduction of the a posteriori RMS from 5 to 3 mm.

determining a dynamic orbit over 8 days (days 010-017/1999) is given. The numbers in Table 2 confirm that kinematic orbits can be used for gravity field determination giving results comparable to the CHAMP models EIGEN-1S and EIGEN2, see Reigber et al. (2003). TUM-1S is the first gravity model derived from purely kinematic orbits and the energy integral.

The quality of the CHAMP GPS data is much better than that of JASON-1 (many data gaps per day, L2-ramps, etc.). This is the reason why CHAMP orbits in Table 2 seem more accurate than JASON-1 orbits. In addition, we should keep in mind that the EIGEN2 and the TUM-1S models were derived from CHAMP SST data only and we may therefore expect a better performance for CHAMP. Such a tuning effect (ICESAT satellite) seems to be less pronounced for the new GRACE gravity model, see Rim et al. (2003).

4.1 GRACE kinematic baseline in space with ambiguity resolution

Let us now see what accuracy might be achievable for the inter-satellite baseline between the two GRACE satellites using a kinematic approach. In order to do this, phase zero-difference measurements were simulated for both GRACE satellites, assuming the noise level and the number of GPS satellite tracked to be similar to CHAMP (only a noise of 1.1 mm was considered with multipath included in this noise level). A typical noise value for the a posteriori RMS of the phase zero-differences in CHAMP kinematic POD is about 1.5–2.0 mm or 1.2–1.4 mm when using double-differences. Whereas zero-differences are mainly affected by the GPS satellite orbit/clock errors, double-differences primarily reflect ground station specific errors like troposphere, multipath, etc. Therefore, the noise level of 1.1 mm adopted for the GRACE simulation might be considered rather pessimistic, keeping in mind that for the short GRACE baseline (220 km), the effect of GPS orbit errors should be only about 0.2 mm, tropospheric refraction is nonexistent and multipath is expected to be very small.

Fig. 6 shows GRACE kinematic baseline results with float, Fig. 7 those with fixed ambiguities. In both cases, GRACE-B was kept fixed to the a priori orbit and GRACE-A positions were estimated kinematically. Comparing these two figures, one can clearly notice that ambiguity resolution de-correlates kinematic coordinates and ambiguities and changes

the colored noise present in the kinematic positions of the float solution into white noise. A decrease of the a posteriori RMS from 5 to 3 mm for the along-track component can also be noticed. Ambiguity resolution was performed as explained in Section 2.1 (Melbourne-Wübbena wide-laning, narrow-lane bootstrapping) and all ambiguities were correctly resolved. Similar results are to be expected with real GRACE data. GRACE GPS data will be a very nice playground to study, for the first time, an inter-satellite baseline with the unique possibility to validate the results with the much more accurate measurements of the K-band link.

4.2 Gravity field determination based on kinematic orbits and clocks in space

As a direct consequence of Einstein's general theory of relativity, a source of radiation in a gravitational potential V_B appears shifted in frequency to an observer in a different gravitational potential V_A by an amount $\Delta f/f = -\Delta V/c^2$, where $\Delta V = V_B - V_A$ is the gravitational potential difference between the source B and the observer A positions.

If we approximate the gravitational potential by the central term GM/r (radial geocentric distance r, gravity constant GM, and speed of light in vacuum c), we get

$$\frac{\Delta f}{f} = \frac{\Delta V}{c^2} \approx \frac{GM}{c^2 r^2}\Delta r. \qquad (6)$$

If we now assume that two extremely performing clocks in space are stable at the level of 10^{-18} over 15 s (corresponding to $\approx$ 100 km in the orbit of the clocks) and that the gravitational frequency shift between these two clocks can be measured with a similar accuracy, we will be able to directly measure differences in the gravity potential that correspond to a change Δr in the equipotential surface of $\approx$ 1 cm over 100 km. Since kinematic positions can already be determined with an accuracy of 1-3 cm and the relative accuracy between successive epochs is even better, the positions (geometry) of the pair of clocks is well-enough known to support such measurements of the gravity potential difference.

Ultra-stable clocks, matter-wave interferometers and atom lasers based on Bose-Einstein condensation are developing rapidly and it is now conceivable to fly such a clock aboard the International Space Station (ACES mission), see Salomon et al. (2001). Space offers weightlessness and atoms can be cooled to such low temperatures that the Earth gravity field represents a major perturbation to their motion. Micro-gravity conditions aboard the Space Station allow to keep these atoms in the observation volume for several seconds (Salomon et al. (2001)), much longer than on the ground, which leads to the increased stability and accuracy. Although a frequency stability of 10^{-16}–10^{-17} over one day still does not meet the above requirements for the gravity field determination, the latest developments in high-precision optical spectroscopy outperform today's state-of-the-art caesium clocks, especially over short periods (Udem et al. (2001)).

5 Conclusions

Using CHAMP GPS measurements we showed that it is possible to estimate the orbit of a Low Earth Orbiting (LEO) satellite purely kinematically with the same level of accuracy ($\approx$ 1-3 cm, w.r.t. SLR) as by the widely employed (reduced-)dynamic approaches.

Kinematic precise orbit determination (POD) is a purely geometrical approach and therefore independent of satellite dynamics (e.g. gravity field, air-drag, etc.) and orbit characteristics (e.g. orbit height, eccentricity, etc.).

Kinematic and reduced-dynamic POD for CHAMP was carried out for GPS week 1175 using zero- and double-differences with ambiguity resolution. It has been shown that all developed kinematic and reduced-dynamic approaches allow for a similar accuracy of 1–3 cm (w.r.t. SLR) and that ambiguity resolution changes the orbit by $\approx$ 1–2 cm.

With a new type of linear combination, called LP (based on phase and P-code GPS measurements from only one frequency) we showed that reduced-dynamic LEO POD can be performed with an accuracy of $\approx$ 10 cm.

With simulated data, the baseline between the two GRACE satellites was estimated kinematically and a clear increase in accuracy could be demonstrated by performing ambiguity resolution.

Finally, kinematic POD in combination with space (optical) atomic clocks for gravity field determination was discussed. Space offers weightlessness and atoms can be cooled to such low temperatures that the Earth gravity field represents a major perturbation to their motion. It is to be expected in the near future that the idea of measuring gravitational frequency shift and thus gravitational potential differences between space clocks, together with geometry known from kinematic POD, will become reality.

Acknowledgements. We are grateful to GFZ Potsdam for providing GPS measurements from CHAMP satellite.

References

Balmino G, Perosanz F, Rummel R, Sneeuw N, Sünkel H, 1999. CHAMP, GRACE and GOCE: mission concepts and simulations. Bolletino di Geofisica Teorica ed Applicata. Vol. 40, 309-319.

Bertiger W, Bar-Sever Y, Christensen E, Davis E, Guinn J, Haines B, Ibanez-Meier R, Jee J, Lichten S, Melbourne W, Muellerschoen R, Munson T, Vigue Y, Wu S, Yunck T, Schutz B, Abusali P, Rim H, Watkins M, Willis P, 1994. GPS precise tracking of TOPEX/POSEIDON: results and implication. J Geophys Res 99, C12: 24449-24464.

Bertiger W, Wu SC, 1996. Single frequency GPS orbit determination for low earth orbiters, ION National Technical Meeting, Jan., 1996.

Byun SH, 2003. Satellite orbit determination using triple-differene GPS carrier phase in pure kinematic mode, J. of Geodesy 76: 569-585.

Bock H, 2003. Effiient Methods for Determining Precise Orbits of Low Earth Orbiters Using the Global Positioning System. PhD thesis. Astronomial Institute University of Berne, Berne, Schwitzerland.

Colombo OL, Luthcke SB, Rowlands DD, Chin DS, Poulouse S, 2002. Filtering Errors in LEO Trajectories Obtained by Kinematic GPS with Floating Ambiguities. Presented at The ION Symposium "GPS 2002", Sept. 24-27, 2002, Portland Oregon.

Földváry L, Gerlach Ch, Švehla D, Frommknecht B, Gruber Th, Peters Th, Rothacher M, Rummel R, Sneeuw N, Steigenberger P, 2003. Determination of the Gravity Field From CHAMP Measurements Considering the Energy Integral. IUGG 2003, General Assembly, June 30 - July 11 2003, Sapporo, Japan.

Gerlach Ch, Földváry L, Švehla D, Gruber Th, Wermuth M, Rothacher M, Rummel R, Sneeuw N, Frommknecht B, Peters Th, Steigenberger P, 2003. A CHAMP only Gravity Field Model From Kinematic Orbits Using the Energy Integral. Geoph. Res. Letters, 30(20), 2037, doi:10.1029/2003GL018025.

Melbourne WG, 1985. The Case for Ranging in GPS Based Geodetic Systems. First International Symposium on Precise Positioning with the GPS, C. Goad (Ed), U.S. Department of Commerce, Rockville, Maryland, pp. 373-386.

Reigber Ch, Schwintzer P, Neumayer K-H, Barthelmes F, König R, Förste Ch, Balmino G, Biancale R, Lemoine J-M, Loyer S, Bruinsma S, Perosanz F, Fayard T, 2003. The CHAMP-only Earth Gravity Field Model EIGEN-2. Adv. in Space Research, 31(8), 1883-1888, 2003, doi: 10.1016/S0273–1177(03)00162-5.

Rim HJ, Ries JC, Yoon S, Webb CE, Bettadpur S, Schutz BE, 2003. Gravity model development for ICESAT precision orbit determinatio. IUGG 2003, General Assembly, June 30 - July 11 2003, Sapporo, Japan.

Salomon C, Dimarcq N, Abgrall M, Clairon A, Laurent P, Lemonde P, Santarelli G, Uhrich P, Bernier LG, Busca G, Jornod A, Thomann P, Samain E, Wolf P, Gonzalez F, Guillemot P, Léon S, Nouel F, Sirmain C, Feltham S, 2001. Cold Atoms in Space and Atomic Clocks: ACES. C.R. Acad.Sci. Paris, T2 Série 4, 1313.

Švehla D, Rothacher M, 2002a. Kinematic Orbit Determination of LEOs Based on Zero- or Double-Difference Algorithms Using Simulated and Real SST Data. IAG 2001 Scientific Assembly, Budapest, Vistas for Geodesy in the New Millenium, Adam J, Schwarz KP (Eds), Springer IAG Vol. 125, pp. 322-328.

Švehla D, Rothacher M, 2003b. CHAMP double-difference kinematic orbit with ambiguity resolution. 1st CHAMP Science Meeting, Potsdam, Germany. First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies, Reigber Ch, Lühr H, Schwintzer P (Eds), Springer, pp. 70-77.

Švehla D, Rothacher M, 2003c. Kinematic and Reduced-Dynamic Precise Orbit Determination of Low Earth Orbiters. EGSXXVII General Assembly 2002, Nice, France. Advances in Geosciences 1, 47-56. (http://www.copernicus.org/EGU/adgeo/published_papers.htm)

Švehla D, Rothacher M, 2003d. Kinematic and Reduced-Dynamic Precise Orbit Determination of CHAMP satellite over one year using zero-differences. Poster presented at EGS-AGU-EUG Joint Assembly 06-11 April 2003, Nice, France. (http://www.iapg.bv.tum.de/cstg/leoliterature.html)

Tapley BD, Ries JC, Davis GW, Eanes RJ, Schutz BE, Shum CK, Watkins MM, Marshall JA, Nerem RS, Putney BH, Klosko SM, Luthcke SB, Pavlis D, Williamson RG, Zelensky NP, 1994. Precision orbit determination for TOPEX/POSEIDON. J Geophys Res 99, C12: 24383-24404.

Udem Th, Holzwarth R, Hänsh TW, 2002. Optical frequency metrology. Nature, Vol. 416, 233-237.

Yunck T, Bertiger W, Wu S, Bar-Sever Y, Christensen E, Haines B, Lichten S, Muellerschoen R, Vigue Y, Willis P, 1994. First assessment of GPS-based reduced dynamic orbit determination on TOPEX/POSEIDON. Geophys Res Letters 21: 541-544.

Gravity Field Recovery and Validation by Analysis of Short Arcs of a Satellite-to-Satellite Tracking Experiment as CHAMP and GRACE

K.H. Ilk, M. Feuchtinger, T. Mayer-Gürr
Institute of Theoretical Geodesy, University of Bonn, Nussallee 17, D-53115 Bonn, Germany

Abstract. A procedure for gravity field determination is presented based on the analysis of short arcs of low flying satellites as CHAMP and GRACE. The analysis technique can be applied to high-low as well as to low-low satellite-to-satellite tracking in a consistent way. The method allows to recover the global gravity field of the Earth, combined with a regional gravity field refinement in regions with rough gravity field features. To detect regional residual signals in the observations a pre-processing step is being applied based on the energy balances along the short arcs. This procedure will be applied as well as the validation of the regional gravity field refinement in a post-processing step. First results are presented by using Post-processed Science Orbits (PSO) of CHAMP for a region with rough gravity field features. To demonstrate the applicability to a tailored low-low satellite-to-satellite tracking experiment the procedure is applied to a GRACE simulation scenario.

Keywords: Gravity field modeling – Regional gravity field recovery – Energy balance relations – Jacobi integral – CHAMP – GRACE – Satellite-to-Satellite-tracking

1 Introduction

The space borne gravity field missions CHAMP and GRACE will improve our knowledge of the Earth's gravity field especially in the low, medium and short wavelengths ranges. Subsequent solutions by using the observations collected over a period of time of, e.g., one month, should enable to derive time dependencies of the gravity field parameters. The temporal effects of the gravity field are very small and most of them are of regional character. So the question arises whether global solutions based on a series of spherical harmonics are the adequate model to represent the static and temporal part of the gravity field for all tasks of these missions. It seems to be useful to model the gravity field in regions with rough gravity field features by base function with local support. Closely related to the problem of gravity field modeling is the question of applying an adequate recovery procedure which takes account of the inherent nature of satellite-to-satellite tracking observables. While satellite gravity gradiometry measures directly point-wise the inner geometry of the gravity field, the classical orbit perturbation methods, e.g., based on Kaula's perturbation theory or by differential orbital improvement techniques analyze accumulated orbit perturbation effects of rather long arcs. From its intrinsic gravity field detection point of view satellite-to-satellite tracking is somewhere in-between point measurements and accumulated orbit perturbations. It seems to be a natural choice to select satellite arcs, which are not too long, to keep the effects of disturbing forces as low as possible.

The solution strategy presented here allows the global recovery of the gravity field combined with regional refinements in geographical areas with rough gravity field features in a consistent way. The physical model is based on Newton's equation of motion applied to short arcs passing over the selected geographical region and formulated as a boundary value problem of a corresponding integral equation of Fredholm type. The integrands contain the reference and residual gravity fields and specific disturbing forces. The reference gravity field representing the low and medium frequency gravity field features is expressed by a series of spherical harmonics complete up to an appropriate degree, while the regional parts are represented by space localizing base functions.

Especially for a regional refinement of the gravity field it is important to validate the absolute/relative orbits together with the gravity field parameters. This can be performed in a two-step procedure. In a pre-processing step, the reference spherical harmonics representation of the gravity field is being validated together with the orbits derived from a precise kinematical or reduced-dynamic orbit determination procedure. In a post-processing step, the regionally refined gravity field is validated again by using the absolute/relative orbits. The validation procedure is based on the computation of an extended Jacobi integral along the satellite orbits. The pre-processing step uncovers deficiencies in the orbits and the reference gravity field representation simultaneously. The post-processing step controls the refinement of the regional gravity field after the regional recovery.

2 Setup of the mathematical model

The mathematical model of the gravity field analysis technique is based on Newton's equation of mo-

tion, with the position vector $\mathbf{R}$ to the satellite, in Fig. 1, e.g., $\mathbf{R}_1$ or $\mathbf{R}_2$, referred to a geocentric (quasi-) inertial system, and $\mathbf{x}$ the column matrix of unknown field parameters, either potential coefficients or any sort of local gravity field parameters (Schneider, 1967),

$$\ddot{\mathbf{R}}(t) = \mathbf{a}(t;\mathbf{R},\dot{\mathbf{R}};\mathbf{x}) . \tag{1}$$

The formulation of the equation of motion as a boundary value problem reads,

$$\mathbf{R}(\tau) = (1-\tau)\mathbf{R}_A + \tau\mathbf{R}_B - T^2 \int_{\tau'=0}^{1} K(\tau,\tau')\, \mathbf{a}(\tau';\mathbf{R},\dot{\mathbf{R}},\mathbf{x})\, d\tau', \tag{2}$$

with the normalized time variable τ,

$$\tau = \frac{t-t_A}{T} \quad \text{with} \quad t \in [t_A, t_B], \quad T = t_B - t_A , \tag{3}$$

the integral kernel

$$K(\tau,\tau') = \begin{cases} \tau(1-\tau'), & \tau \le \tau', \\ \tau'(1-\tau), & \tau' \le \tau, \end{cases} \tag{4}$$

and the boundary values,

$$\mathbf{R}_A := \mathbf{R}(t_A), \quad \mathbf{R}_B := \mathbf{R}(t_B), \quad t_A < t_B . \tag{5}$$

The solution can be given in the form as follows,

$$\mathbf{R}(t) = (1-\tau)\mathbf{R}_A + \tau\mathbf{R}_B + \sum_{\nu=1}^{\infty} \mathbf{R}_\nu \sin(\nu\pi\tau) . \tag{6}$$

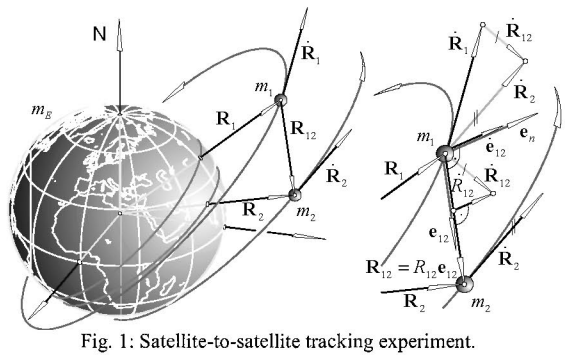


Fig. 1: Satellite-to-satellite tracking experiment.

The coefficients $\mathbf{R}_\nu, \nu = 1,2,...,\infty$, of formula (6) can be derived from the right-hand side of equation (1) (for details see Ilk et al., 1995)

$$\mathbf{R}_\nu = -\frac{2T^2}{\pi^2\nu^2} \int_{\tau'=0}^{1} \sin(\nu\pi\tau')\, \mathbf{a}(\tau';\mathbf{R},\dot{\mathbf{R}},\mathbf{x})\, d\tau' , \tag{7}$$

or by measurements, as in case of accelerations,

$$\mathbf{R}_\nu = -\frac{2T^2}{\pi^2\nu^2} \int_{\tau'=0}^{1} \sin(\nu\pi\tau')\, \ddot{\mathbf{R}}(\tau')\, d\tau' . \tag{8}$$

This formula can be transformed by partial integration,

$$\mathbf{R}_\nu = \frac{2T}{\pi\nu} \int_{\tau'=0}^{1} \cos(\nu\pi\tau')\, \dot{\mathbf{R}}(\tau')\, d\tau' , \tag{9}$$

or by another partial integration as,

$$\mathbf{R}_\nu = 2 \int_{\tau'=0}^{1} \sin(\nu\pi\tau') \left(\mathbf{R}(\tau') - (1-\tau')\mathbf{R}_A - \tau'\mathbf{R}_B\right) d\tau' . \tag{10}$$

The acceleration function

$$\mathbf{a}(\tau';\mathbf{R},\dot{\mathbf{R}},\mathbf{x}) = \mathbf{a}_d(\tau') + \nabla V(\tau';\mathbf{R};\mathbf{x}_0) + \nabla T(\tau';\mathbf{R};\Delta\mathbf{x}) , \tag{11}$$

can be separated in a disturbance part $\mathbf{a}_d$, which represents the non-conservative disturbing forces, in a reference part ∇V, representing the global gravity field features by the reference potential coefficients $\mathbf{x}_0$ and in an anomalous part ∇T, modeling the corrections $\Delta\mathbf{x}$ to the global gravity field parameters.

A first version of the mathematical model is given by equation (2) together with the acceleration function (11). The geocentric positions $\mathbf{R}(t)$ of the arcs over the analysis area represent the observations. The unknowns are the corrections to the field parameters, $\Delta\mathbf{x}$. These are, in case of a global gravity field recovery, corrections to the coefficients of a spherical harmonics expansion of the gravitational potential V or, in case of a regional recovery, the parameters of space localizing base functions modeling the anomalous potential T. Besides this space domain method a spectral domain version is possible, based on formula (7) together with the acceleration function model (11). The amplitudes $\mathbf{R}_\nu$ with $\nu = 1,2,...,\nu_{\max}$, are determined from the observations by numerical quadrature of equations (8) to (10), respectively, depending on the observation type, or by a prior least squares adjustment fit to the observations. These model alternatives can be used in case of the high-low satellite-to-satellite configuration, if the orbits are determined precisely by kinematical orbit determination methods, based on precise GPS measurements.

If precise inter-satellite functionals as line-of-sight ranges or range-rate measurements are available, as in case of the GRACE mission, the mathematical model can be based on the equation of motion for the line-of-sight distance (Fig. 1),

$$\ddot{R}_{12}(t) = \frac{\dot{\mathbf{R}}_{12}^2 - \dot{R}_{12}^2}{R_{12}} + \mathbf{e}_{12} \cdot \mathbf{g}(t;\mathbf{R}_{12},\mathbf{R}_1,\dot{\mathbf{R}}_1,\dot{\mathbf{R}}_2;\mathbf{x}) , \tag{12}$$

again formulated as a boundary value problem,

$$R_{12}(t) = (1-\tau)R_{12,A} + \tau R_{12,B} - T^2 \int_{\tau'=0}^{1} K(\tau,\tau') \cdot \left(\frac{\dot{\mathbf{R}}_{12}^2 - \dot{R}_{12}^2}{R_{12}} + \mathbf{e}_{12} \cdot \mathbf{g}(\tau';\mathbf{R}_{12},\mathbf{R}_1,\dot{\mathbf{R}}_1,\dot{\mathbf{R}}_2;\mathbf{x}) \right) d\tau', \tag{13}$$

The solution can be given in an analogous way as before by

$$R_{12}(t) = (1-\tau)R_{12,A} + \tau R_{12,B} + \sum_{\nu=1}^{\infty} R_{12,\nu} \sin(\nu\pi\tau) , \tag{14}$$

with the coefficients $R_{12,\nu}, \nu = 1,2,...,\infty$,

$$R_{12,\nu} = -\frac{2T^2}{\pi^2\nu^2} \int_{\tau'=0}^{1} \sin(\nu\pi\tau') \cdot \left(\frac{\dot{\mathbf{R}}_{12}^2 - \dot{R}_{12}^2}{R_{12}} + \mathbf{e}_{12} \cdot \mathbf{g}(\tau';\mathbf{R}_{12},\mathbf{R}_1,\dot{\mathbf{R}}_1,\dot{\mathbf{R}}_2;\mathbf{x}) \right) d\tau'. \tag{15}$$

The acceleration function,

$$\mathbf{g}(\tau';\mathbf{R}_{12},\mathbf{R}_1,\dot{\mathbf{R}}_1,\dot{\mathbf{R}}_2;\mathbf{x}) = \mathbf{g}_d(\tau') + \nabla V_{(12)E}(\tau';\mathbf{R}_{12},\mathbf{R}_1;\mathbf{x}_0) + \nabla T_{(12)E}(\tau';\mathbf{R}_{12},\mathbf{R}_1;\Delta\mathbf{x}). \tag{16}$$

can be separated again in a disturbance part $\mathbf{g}_d$, in a

reference part $\nabla V_{(12)E}$, now modeled by the reference tidal potential of the Earth,

$$\nabla V_{(12)E}(\tau';\mathbf{R}_{12},\mathbf{R}_1;\mathbf{x}_0)=\nabla\big(V(\mathbf{R}_1+\mathbf{R}_{12})-V(\mathbf{R}_1)\big), \quad (17)$$

representing the global gravity field features and in an anomalous part $\nabla T_{(12)E}$,

$$\nabla T_{(12)E}(\tau';\mathbf{R}_{12},\mathbf{R}_1;\Delta\mathbf{x})=\nabla\big(T(\mathbf{R}_1+\mathbf{R}_{12})-T(\mathbf{R}_1)\big), \quad (18)$$

which is again parameterized either by corrections to the global gravity field parameters or by parameters of a linear approximation with space localizing base functions. The coefficients $R_{12,\nu}, \nu=1,2,...,\infty$, of formula (14) can be derived by inter-satellite measurements of different types, e.g., in case of relative accelerations

$$R_{12,\nu}=-\frac{2T^2}{\pi^2\nu^2}\int_{\tau'=0}^{1}\sin(\nu\pi\tau')\,\ddot{R}_{12}(\tau')\,d\tau', \quad (19)$$

and/or in case of range-rate measurements,

$$R_{12,\nu}=\frac{2T}{\pi\nu}\int_{\tau'=0}^{1}\cos(\nu\pi\tau')\,\dot{R}_{12}(\tau')\,d\tau', \quad (20)$$

and/or in case of inter-satellite range observations,

$$R_{12,\nu}=2\int_{\tau'=0}^{1}\sin(\nu\pi\tau')\cdot \cdot\big(R_{12}(\tau')-(1-\tau')R_{12,A}-\tau'R_{12,B}\big)\,d\tau', \quad (21)$$

respectively. Again, a space domain model based on equation (13) with the acceleration function according to (16) or a spectral domain model based on formula (15) with the spectral observations due to equations (19) to (21) can be formulated. For both alternatives the normal equations are established and solved by a regularized solver of Tichonov type, where the regularization parameter is preferably computed according to the variance component estimation procedure due to Koch and Kusche (2002).

3 Gravity field representation

The gravity field is divided as usual in a reference part $V(P)$ and in a disturbing part $T(P)$ represented by a linear approximation in terms of space localizing base functions,

$$T(\mathbf{R}(P))=T(P)=\sum_{i=1}^{I}a_iK(P,Q_i), \quad (22)$$

with the field parameters a_i arranged in a column matrix $\Delta\mathbf{x}:=(a_i, i=1,...,I)^T$ and the base functions,

$$K(P,Q_i)=\sum_{n=0}^{\infty}k_n\left(\frac{R_E}{R}\right)^{n+1}P_n(P,Q_i). \quad (23)$$

The coefficients k_n are the degree variances of the gravity field spectrum to be determined,

$$k_n=\sum_{m=0}^{n}\left(\Delta\bar{c}_{nm}^2+\Delta\bar{s}_{nm}^2\right). \quad (24)$$

R_E is the mean equator radius of the Earth, R the distance of a field point from the geo-center and $P_n(P,Q_i)$ are the Legendre's polynomials depending on the spherical distance between a field point P and the nodal points Q_i of the set of base functions. With this definition the base functions $K(P,Q_i)$ can be interpreted as isotropic and homogeneous harmonic spline functions (Freeden et al., 1998). The nodal points Q_i are defined on a grid generated by a uniform subdivision of an icosahedron of twenty equal-area spherical triangles. In this way the global pattern of spline nodal points shows approximately uniform nodal point distances. Any other functional of the disturbing potential can be derived from this representation in a simple way.

4 Validation procedure

The validation procedure shall detect inconsistencies in the precisely determined orbits and the reference gravity field model in the pre-processing step on the one hand, and shall control the refinement of the regional gravity field in the post-processing step on the other hand. The reference gravity field is based on a recent global gravity field model or derived by a global application of the recovery techniques as described in the last section. The validation procedure should be based preferably on a different functional model than for the gravity field recovery procedure. A proper tool to validate the consistency of orbit and gravitational field is the computation of the total energy of the satellite along its orbit. The classical energy balance has to be extended for this task to cover also alternative energy contributions caused by a non-uniform Earth rotation or by non-conservative forces as the time dependent tidal forces or surface forces acting on the satellites. This modified energy balance can be expressed by an extended Jacobi integral (Ilk and Löcher, 2003):

$$J_{LEO}:=E'(t)+E'_{akz}|_{t_0}^{t}+E'_{M}|_{t_0}^{t}+E'_{S}|_{t_0}^{t}+E'_{E}|_{t_0}^{t}+ +E'_{Earth\,Tides}|_{t_0}^{t}+E'_{Ocean\,Tides}|_{t_0}^{t}=const. \quad (25)$$

The term $E'(t)$ represents the (specific mass related) "classical" Jacobi integral

$$E'(t)=E'_{kin}(t)+E'_{cen}(t)+E'_{pot}(t)= =\frac{1}{2}\dot{\mathbf{R}}'^2-\frac{1}{2}\big(\boldsymbol{\Omega}\times\mathbf{R}'\big)^2-V'(\mathbf{R}')=const, \quad (26)$$

with the (specific) kinetic energy $E'_{kin}(t)$, the centrifugal energy $E'_{cen}(t)$ and the potential energy $E'_{pot}(t)$. $E'_{akz}|_{t_0}^{t}$ is the energy contribution by the measured accelerations, considering the surface forces like air drag and solar pressure (including Earth albedo). The terms $E'_{M}|_{t_0}^{t}$ and $E'_{S}|_{t_0}^{t}$ represent the energy contributions from the direct tidal forces of Moon and Sun and $E'_{E}|_{t_0}^{t}$ is the contribution from the Eulerian accelerations due to a non-constant rotation of the Earth. The terms $E'_{Earth\,Tides}|_{t_0}^{t}$ and $E'_{Ocean\,Tides}|_{t_0}^{t}$ indicate the energy contributions from the solid Earth and ocean tides. For the relative motions within an SST configuration an analog ex-

tended Jacobi integral has to be formulated for the GPS satellites. The energy contributions from the Earth and ocean tides to the total energy budget of the satellites of the GPS are small and can be neglected. Instead of the energy contribution from the measured accelerations that from solar radiation pressure, $E'_{rad}|_{t_0}^{t}$, has to be considered; air drag is not relevant at the altitude of these satellites. Therefore, the extended Jacobi integral for a GPS satellite n reads,

$$J_{GPSn} := E'(t) + E'_M|_{t_0}^{t} + E'_S|_{t_0}^{t} + E'_E|_{t_0}^{t} + E'_{rad}|_{t_0}^{t} = const\,. \qquad (27)$$

The energy contributions $E'_{K_i}|_{t_0}^{t}$ of an arbitrary force $\mathbf{K}'$ can be derived by an appropriate formula for numerical quadrature (Löcher, unpub. manuscript),

$$E'_{K_i}|_{t_0}^{t} = -\int_{t_0}^{t} \mathbf{K}'_i \cdot d\mathbf{R}' = -\int_{t_0}^{t} \mathbf{K}'_i \cdot \left(\dot{\mathbf{R}} - \boldsymbol{\Omega}\times\mathbf{R}'\right) dt\,. \qquad (28)$$

For example, the specific (direct) tidal forces $\mathbf{G}'_{(Sat,E)i}$ exerted by the Moon or the Sun (mass M_i, $i = M,S$) onto the satellite's motion referred to a geocentric reference system result in an energy contribution

$$E'_i|_{t_0}^{t} = -\int_{t_0}^{t} \mathbf{G}'_{(Sat,E)i} \cdot \left(\dot{\mathbf{R}} - \boldsymbol{\Omega}\times\mathbf{R}'\right) dt\,. \qquad (29)$$

Those quantities which refer to the rotating reference system with the time variable rotation vector $\boldsymbol{\Omega}$ are marked by an apostrophe. The time dependency of the rotation vector $\boldsymbol{\Omega}$ results in an energy contribution

$$E'_E|_{t_0}^{t} = -\int_{t_0}^{t} \mathbf{E}' \cdot \dot{\mathbf{R}}\, dt = \int_{t_0}^{t} \dot{\boldsymbol{\Omega}} \cdot \left(\mathbf{R}\times\dot{\mathbf{R}}\right) dt\,, \qquad (30)$$

caused by the Eulerian acceleration,

$$\mathbf{E}' = -\dot{\boldsymbol{\Omega}}'\times\mathbf{R}' = -\dot{\boldsymbol{\Omega}}\times\mathbf{R}\,. \qquad (31)$$

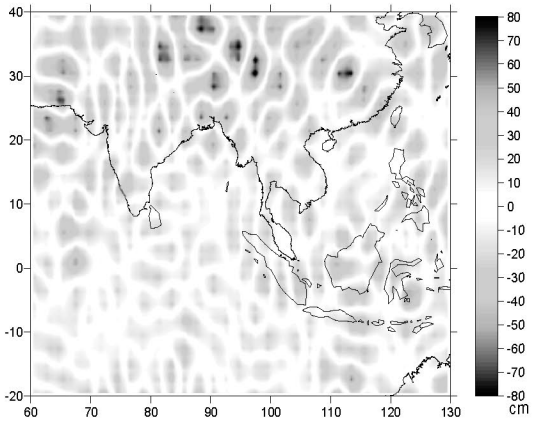


Fig. 2: Residuals of the combined GRACE1 solution.

As mentioned before, deviations from the Jacobi constants uncover inconsistencies of orbits and force function models. But it is not easy to detect the sources of these inconsistencies. Further investigations are necessary to come up with a careful analysis of these inconsistencies. It would be very helpful to separate space-dependent and time-dependent effects. Nevertheless, the comparison of the time evolution of the Jacobi constant before the regional refinement and afterwards are very helpful to validate the recovery result.

5 Simulation scenario for GRACE

Unfortunately, real data from the GRACE mission are not available to the public yet. Therefore, a simulation scenario has been generated to investigate the regional recovery procedure. The nearly circular orbits of the two GRACE satellites have been computed at an altitude of 490km, following each other in a distance of approximately 230km. The pseudo-real gravity field has been simulated by a spherical harmonics expansion complete up to degree 300 (EGM96). The satellite positions of both satellites have been modified by white noise with an rms of 2cm. The range and range-rate measurements have been generated every 5sec covering a 30days mission period and corrupted by white noise with an rms of 10 μm and 5 μm/sec, respectively. The recovery area is shown in Fig. 2; a stripe of 15° has been added around this area to avoid geographical truncation effects. The number of orbits passing over this region within the 30-days analysis period amounts to 237. The number of gravity field parameters modeling the regional gravity field refinement corresponds to the number of grid points adapted to the icosahedral triangles. This number depends on the resolution envisaged, in the present case 6969 nodal points were used, resulting in a mean nodal point distance of about 120km for the range-rate SST configuration. The regional recovery can be based on the analysis of the GRACE orbits based on model (7) with observations of type (10) for both GRACE satellites separately (about 48870 observations) and based on a combined solution. The 3137 nodal points with a mean distance of about 220km represent in this case a significantly lower resolution than in the SST-case. The differences to the pseudo-real gravity field show a maximal error of 121cm, an average deviation of 26cm and an rms of 33cm in terms of geoid heights for GRACE1. The analog numbers for GRACE2 are 122cm, 22cm and 29cm and for the combined solution 111cm, 21cm and 28cm. The residuals to the true solution are shown in Fig. 2. The low-low inter-satellite observations are treated according to model (13) with observations of type (20) or (21). The total number of pseudo-observations in this case is 21330. The recovery result using range-rates (formula (20)) with an rms of 5μm/s shows a maximal error of 30cm, an average deviation from the pseudo-real field of 4.9cm and an rms of 6.1 cm in terms of geoid heights. The corresponding

numbers for ranges (formula (21)) with an rms of 10μm are 84cm , 14.4cm and 18.4cm .

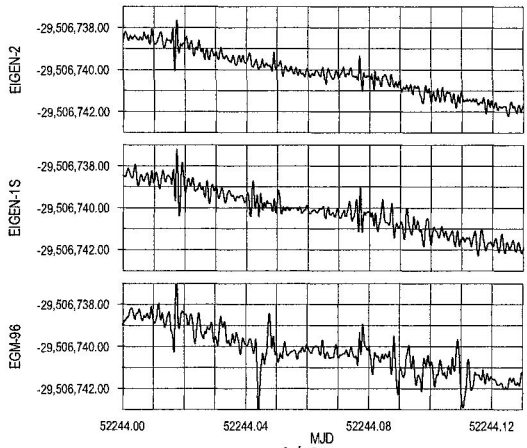


Fig. 3: Energy contributions (m^2/s^2) to the extended Jacobi integral for a two-revolution PSO arc of CHAMP in case of various gravity field models.

6 Application to PSO data of CHAMP

6.1 Data set

To demonstrate the recovery and validation procedure, so-called Post-processed Science Orbits (PSO) for CHAMP (Förste et al., 2001) have been used for a one-month mission period (December 2001). A special feature of the CHAMP mission is the fact that all surface forces are being measured by a three-dimensional accelerometer. These data sets as well as observations and orbits at different production levels are provided by the CHAMP Information System and Data Centre (ISDC). The orbits of the CHAMP data base are referred to a terrestrial reference frame (Lühr et al., 2001). The transformations between the terrestrial and celestial reference frames followed the IERS conventions. The computation of the tidal energy requires the ephemeris of Moon and Sun referred to the ITRF96. For this task the numerical ephemeris DE405 of the Jet Propulsion Laboratory (JPL) have been used together with the software to perform the processing of these data. The accelerometer measurements to determine the surface force energy for the CHAMP orbit have been processed according to the rules of the CHAMP data formats (Förste et al., 2001) supplied by (Reigber, 2002). Finally, the energy contributions caused by the Earth tides as well as by the ocean tides have been based on the models published by McCarthy (1996).

6.2 A-priori validation

The a-priori validation step provides information on the consistency of the global gravity field and the orbits used for the analysis. In case of inconsistencies the Jacobi constant shows systematic effects which may be caused by an insufficient modeling of the orbits. Rough gravity field features which are not covered by the set of potential coefficients can generate such systematic effects in the temporal evolution of the Jacobi constant as outlined in section 4. The twenty-minutes arc in Fig. 5 clearly shows systematic deviations of the Jacobi integral from a constant which gives a hint to model deficiencies in the global gravity field. As pointed out, there are no reliable criteria yet to decide which contributions are caused by orbit determination errors or by inconsistencies in the force function models, especially in the gravity field model. Further investigations are necessary at this point.

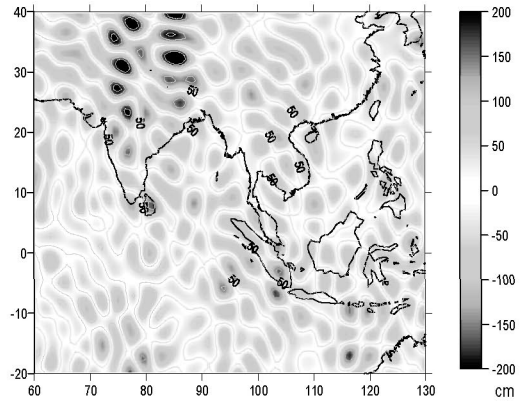


Fig. 4: Regional refinement of the CHAMP solution based on EIGEN-2 and a mission period of 30 days.

In any case, because of the inhomogeneous structure of the gravity field there are good arguments for the assumption that rough gravity field structures in specific regions are mapped onto the orbits and can be analyzed for a regional gravity field improvement. These are the cases where a focus-in procedure becomes justified. Fig. 3 demonstrates this fact; The series of gravity field models EGM-96, EIGEN-1S (Reigber et al., 2002) and EIGEN-2 (Reigber et al., 2003) show a significant improvement from one model to the next in the deviation from a constant in case of a two-revolution PSO arc (1.12.2001), if the secular trend, most probably caused by an insufficient calibration of the accelerometer, is being removed. Still, residual signals seem to be present.

6.3 Regional gravity field recovery

For the regional gravity field refinement the recovery technique based on the analysis of short arcs covering a time period of 30 days has been applied according to model (7) with observations of type (10). A total number of 36850 observations has been used for the establishment of the normal equation system. The recovery area is shown in Fig.6; an additional security stripe of 10° around the recovery area has been considered to avoid geographical

truncation effects. The parameterization of the residual field was based on harmonic spline functions as defined in section 3. The mean distance between the base function nodal points amounts to approximately 130 km . The recent GFZ CHAMP–only gravity model EIGEN-2 has been used as reference gravity field (Reigber et al., 2003).

6.4 A-posteriori validation

To demonstrate the post-processing validation procedure, the Jacobi integral has been computed for the 20-minutes arc of Fig. 5 based on EIGEN-2 plus the regional refinement. It shows a significant improvement, despite the fact, that still systematic deviations from a constant can be detected. The validation check has been applied to various other arcs not used for the regional recovery, resulting in a similar tendency.

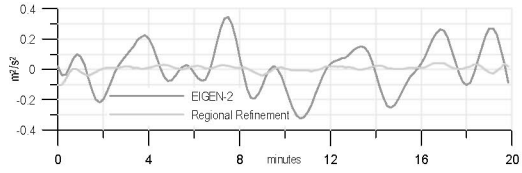


Fig. 5: Jacobi integrals for an arc of 20 minutes not included in the regional recovery refinement.

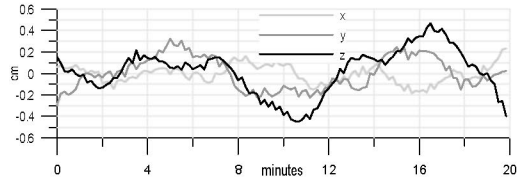


Fig. 6: Differences between an arc based on Eigen-2 and the PSO arc.

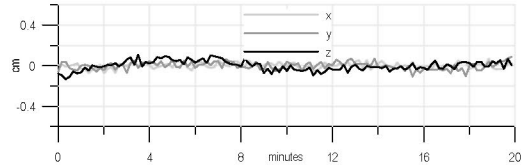


Fig. 7: Differences between an arc based on the regional gravity field refinement and the PSO arc.

These improvements can be seen as well if the predicted ephemeris of this arc are inspected based on the EIGEN-2 gravity field model (Fig. 6) on the one hand and by including the regional refinement (Fig. 7) on the other hand and compared to the PSO arc. Despite the fact that deficiencies in the orbit fits can be seen also by inspecting the orbit residuals the energy quantities may tell more about the character of these residuals and their causes. A typical example is the effect of surface forces which may be seen more clearly by the temporal evolution of the Jacobi integral. The secular trend gives a clear hint to a specific transfer of energy into another energy form.

7 Conclusions

The results of this investigation demonstrate that a regional gravity field refinement is a valuable complement to a global gravity field recovery. Despite the fact that the CHAMP mission is not a proper candidate for a regional focus-in procedure, the example from South-East-Asia clearly shows, that even a high-quality product as Eigen-2 can be improved in specific regions. The recovery results of a simulated GRACE scenario underline the quality of a regional focus of the gravity field. An a-priori check of the temporal evolution of the Jacobi integral may, on the one hand, uncover deficiencies in global models and can, on the other hand, demonstrate the improvements in specific regions by the a-posteriori check.

Acknowledgements:

We are grateful to the GeoForschungsZentrum (GFZ) Potsdam for providing the data for this investigation. The support by BMBF and DFG within the frame of the Geotechnologien Programm is gratefully acknowledged.

References:

Förste, Ch., Schwintzer, P., Reigber, Ch. (2001): The CHAMP Data Format, Internal publication, CH-GFZ-FD-012

Freeden, W., Gervens, T., Schreiner, M. (1998): Constructive Approximation on the Sphere, Oxford University Press, Oxford, 1998

Ilk, K.H., Rummel, R., Thalhammer, M. (1995): Refined Method for the Regional Recovery from GPS/SST and SGG, CIGAR III/2, ESA contract No. 10713/93/F/FL, European Space Agency, 1995

Ilk, K.H., Löcher A. (2003): The Use of Energy Balance Relations for Validation of Gravity Field Models and Orbit Determination Results, presented at the General Assembly of the IUGG 2003, Sapporo, Japan, 2003

Koch, K.-R., Kusche, J. (2002): Regularization of geopotential determination from satellite data by variance components, Journal of Geodesy 76, 5, pp.259-268, 2002

Lühr, H., Grunwaldt, L., Förste, Ch., Schwintzer, P., Reigber, Ch., (2001): CHAMP Reference Systems, Transformations and Standards, Int. publication, CH-GFZ-RS-002

McCarthy, D.D. (ed.), (1996): IERS Conventions 1996, Central Bureau of IERS – Observatoire de Paris, Paris, 1996

Reigber, Ch., (2002) : CHAMP Newsletter No. 9, GFZ, Newsletter_CHAMP_009

Reigber, Ch., Balmino, G., Schwintzer, P., Biancale, R., Bode, A., Lemoine, J.-M., Koenig, R., Loyer, S., Neumayer, H., Marty, J.-C., Barthelmes, F., Perosanz, F., Zhu, S.Y., (2002): A high quality global gravity field model from CHAMP GPS tracking data and Accelerometry (EIGEN-1S), Geophysical Research Letters, 29(14), 10.1029/ 2002GL015064

Reigber, Ch., Schwintzer, P., Neumayer, K.-H., Barthelmes, F., König, R., Förste, Ch., Balmino, G., Biancale, R., Lemoine, J.-M., Loyer, S., Bruinsma, S., Perosanz, F., Fayard, T. (2003): The CHAMP-only Earth Gravity Field Model Eigen-2, Advances in Space Research 31(8), 1883-1888, 2003

Schneider, M. (1967): Lösungsvorschlag zum Bahnbestimmungsproblem, BWF Bericht W67-35, 1967

Combination of Sea-Level Observations and an Oceanographic Model for Geodetic Applications in the Baltic Sea

Kristin Novotny, Gunter Liebsch, Reinhard Dietrich
TU Dresden, Institut für Planetare Geodäsie, D-01062 Dresden, Germany
Andreas Lehmann
Institut für Meereskunde an der Universität Kiel, Düsternbrooker Weg 20, D-24105 Kiel, Germany

Abstract. Satellite altimetry and tide gauges provide observations of the instantaneous sea surface. As a further source an oceanographic model of the Baltic Sea also provides sea-level heights. The different sea-surface heights are of comparable precision and complement each other well in terms of their spatial and temporal resolution and their information content.
When the different sea-level information are combined, variations in the time series can be reduced, the quality of the observations can be checked, and phenomena of special interest can be separated.
One application for the combination of observed and modelled heights is the estimation of secular sea-level changes. By reducing a large part of the high frequency sea-level variations in the observations, the trend estimation from short time series can be improved. This is of interest especially for satellite altimeter observations, which, though having a high spatial resolution, cover a rather short time span.
Keywords. Satellite altimetry, tide gauge, oceanographic model, Baltic Sea.

1 Introduction

Regular sea-level observations in the Baltic Sea began in 1794 at the tide gauge Stockholm (Ekman, 1988), and since then a dense network of tide gauge stations along the coasts of the Baltic Sea has been developed. The long observation time series allow the study of secular changes in the behaviour of the Baltic Sea sea level, and especially the study of the postglacial rebound in Fennoscandia (see e.g. Ekman, 1992; Ekman, 1998; Dietrich and Liebsch, 2000). Due to large sea-level variations particularly in annual and interannual time scales, long time series of at least

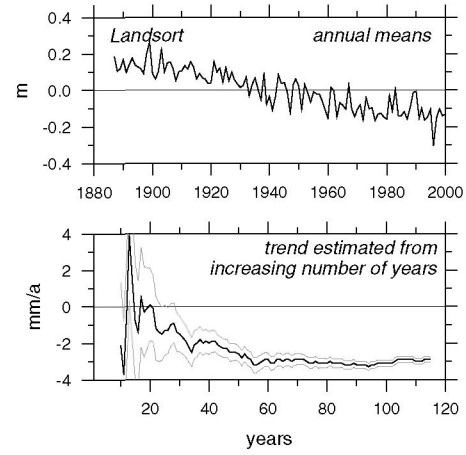


Fig. 1: Annual mean sea level at Landsort, 1887-2000 (top). The bottom plot shows the dependence of the estimated trend values from the length of the time series. Trends were estimated from an increasing number of years, starting 1887. The grey lines indicate the uncertainty (1σ) of the trend estimation.

60 to 70 years length are necessary to determine the secular trend in the observed sea level at a tide gauge station reliably (Figure 1) (Liebsch, 1997).
Satellite altimetry provides consistent observations of the sea surface with a high spatial coverage over the last 10 years. However, the regional trends estimated from these short time series may be biased due to interannual sea-level variations.
This study focusses on the combination of the observations with a precise, high-resolution oceanographic model of the Baltic Sea. The tie between long-term tide gauge observations, the consistent height reference of satellite altimetry and precise modelled sea-surface heights can be useful for different geodetic applications, like e.g. the estimation of sea-level topography, sea-level change or altimeter validation.

Table 1: The Baltic Sea oceanographic model

- Bryan-Cox-Semtner general ocean circulation model with free surface, updated to the Baltic Sea and coupled to a dynamic-thermodynamic sea ice model
- forced by realistic atmospheric conditions (SMHI meteorological data base) and river runoff
- prognostic variables: e.g. oceanic baroclinic current field, 3-D temperature, 2-D surface elevation
- spatial resolution: about 5x5 km (eddy-permitting); 60 vertical levels; temporal resolution: 6 hours; 1979-2002
- western boundary is connected to an artificial North Sea basin; most western boundary adjusted to a constant sea-level height

2 Sea-level variations in the Baltic Sea

The Baltic Sea is a semi-enclosed basin that is connected with the North Sea by narrow and shallow channels at its south western part. This leads to a limited water exchange with the North Sea, and the sea-level variations in the Baltic Sea can be divided into variations due to external and internal effects (Samuelsson and Stigebrandt, 1996).

The water exchange with the North Sea in general affects the fill level of the Baltic Sea, resulting in a nearly uniform sea-level variation signal in the Baltic Sea which depends only on the time, t. That effect can be considered as external ($var_E(t)$). Since the Baltic Sea entrance can be regarded as a low pass filter for sea-level variations that penetrate the Baltic Sea from the North Sea, the external sea-level variation signal in the Baltic Sea is mainly long periodic (Samuelsson and Stigebrandt, 1996).

Effects within the Baltic Sea lead to a different behaviour of the sea surface at different locations, s (internal effect, $var_I(s,t)$). One such internal effect, which is also the most dominant, is the increasing amplitude of the seasonal variation towards the north of the Baltic Sea (Ekman, 1996).

In this study, a precise, high-resolution oceanographic model of the Baltic Sea was utilised (Lehmann, 1995; Lehmann and Hinrichsen, 2000) (Table 1). Among other prognostic variables the model provides 2D surface elevations ($h_{omod}(s,t)$), which show a high consistency with the observed sea-level heights (Novotny et al., 2002a). Due to its boundary condition, which includes a fixed sea-level height at its most western boundary, the externally induced sea-surface variations in the Baltic Sea are not represented in the model (Novotny et al., 2002b). Moreover, by this boundary condition the model admits no long-term sea-level trend. The modelled sea-surface heights, $h_{omod}(s,t)$, reflect mainly the sea-level variations of the Baltic Sea due to internal effects:

$$h_{omod}(s,t) \approx var_I(s,t)$$

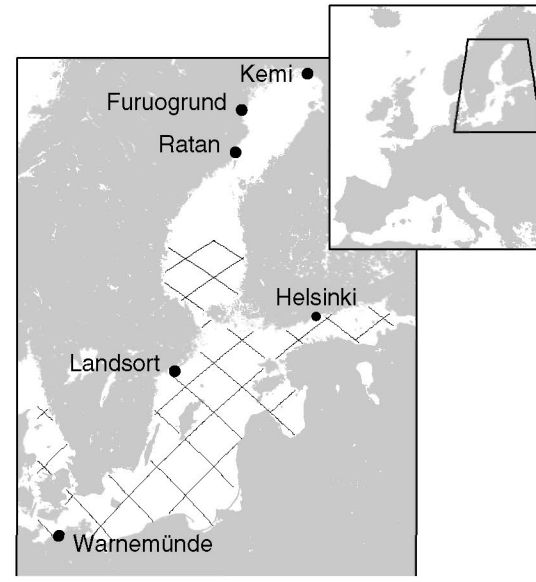


Fig. 2: Baltic Sea region with selected tide gauge stations and used ground tracks of the TOPEX altimeter satellite

The observed sea-level heights contain both the external and internal effects in the sea-surface variability. In addition, the tide gauge data ($h_{tg}(s,t)$) include relative secular changes due to the long-term sea-level change and crustal movements at the tide gauge stations, $T_{rel}(s)$. The altimetric sea-surface heights ($h_{alt}(s,t)$) directly show the geocentric sea-level change, T_{geoc}. While the observed relative trend depends on the location of the tide gauge station, the absolute sea-level change in the Baltic Sea is assumed to be a constant:

$$h_{tg}(s,t) = T_{rel}(s) \cdot t + var_E(t) + var_I(s,t)$$

$$h_{alt}(s,t) = T_{geoc} \cdot t + var_E(t) + var_I(s,t)$$

For this study, monthly mean sea levels at tide gauge stations from the data base of the Permanent Service of Mean Sea Level (PSMSL) were used.

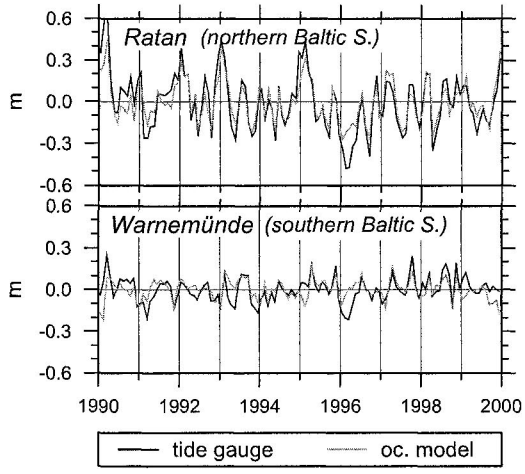


Fig. 3: Comparison of tide gauge observations and modelled sea-surface heights (monthly means) at two stations

Altimetric sea-surface heights from the TOPEX mission (AVISO Corrected Sea Surface Heights; 11/1992-10/2001) were related to the NKG96 geoid (Forsberg et al., 1996). The resulting instantaneous sea-surface topography heights were averaged into monthly mean fill levels of the Baltic Sea, thus averaging out high frequency variations and local effects for individual measurements. Due to data gaps caused by ice coverage during the winter months, observations north of 62° latitude were omitted (Figure 2).

For comparison, the 6 hourly modelled sea-surface heights were also averaged into monthly mean heights.

3 Combination of the sea-level observations with the oceanographic model

Figure 3 shows the comparison of monthly mean sea levels at two tide gauge stations and the corresponding monthly means from the oceanographic model. The increasing sea-level variation towards the north (internal effect) is well reflected by the model. Therefore, the internal effect can be largely eliminated from the observations if corresponding modelled heights are subtracted from the observed sea-level heights.

The effect of the combination of observed and modelled heights is shown in Figure 4. The upper plot again demonstrates the internal sea-level variation effect, $var_I(s,t)$. The reduced time series, that were obtained after the oceanographic model was applied, show a uniform behaviour in terms of their standard deviation (lower plot).

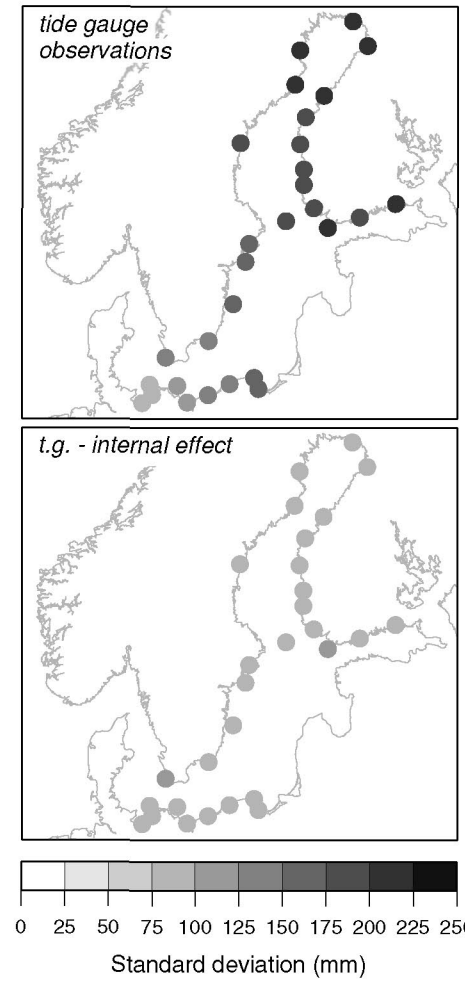


Fig. 4: Standard deviation of the time series of observed sea-level heights at tide gauge stations (top) and of the reduced time series after removing the internal effect in the sea-level variations by applying the oceanographic model (bottom)

In the remaining difference between observed and modelled heights the long-term relative trend at the individual tide gauge station, $T_{rel}(s)$, and the external effect, $var_E(t)$, are left. Comparing such differences at different tide gauge stations, which are well distributed along the Baltic Sea coasts, the spatially uniform behaviour of this external signal can be seen (Figure 5).

Since the dominant part of the external sea-level variation signal is spatially constant, the average of these difference time series can be used as a model of the Baltic Sea sea-level variation due to external effects in the time period considered. A high correlation between this model and the sea-level variations observed at a tide gauge in the

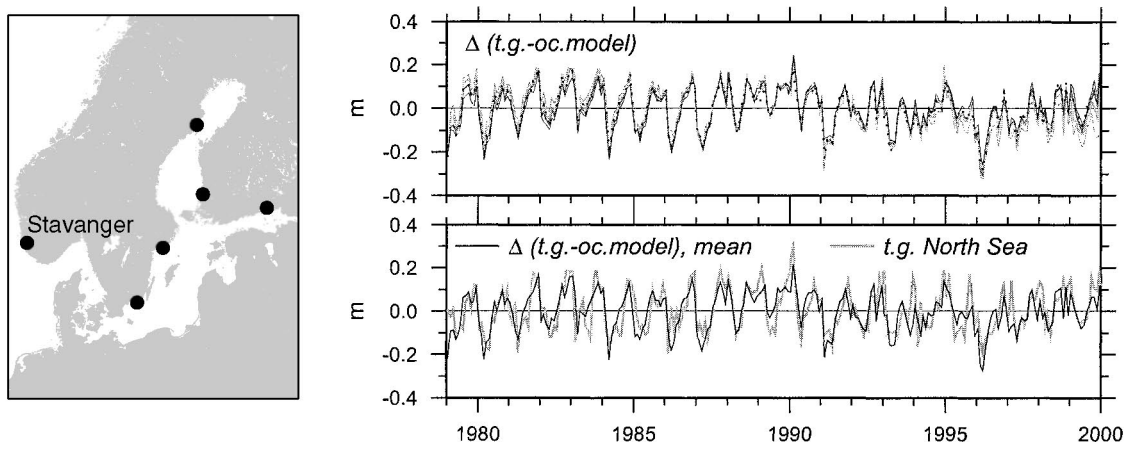


Fig. 5: Differences of observed and modelled sea-surface heights (monthly means) in the Baltic Sea
Top: The individual difference time series at 5 selected stations (here: linear trends removed) show a uniform behaviour, indicating that the external signal is not represented in the Baltic Sea oceanographic model.
Bottom: Averaging the individual difference time series, the external signal can be modelled. It shows a high correlation with sea-level variations observed at a tide gauge in the North Sea (Stavanger).

North Sea is evident, which confirms the external source of the signal (Figure 5).

This modelled external signal can now also be subtracted from the observations, thus removing a large part of the sea-level variations from the observations.

The time series of monthly mean altimetric sea-surface heights can be treated in the same way. Again, the internal effect is directly eliminated by subtracting corresponding modelled heights from the altimetric sea-surface height observations. To account for the remaining external effect, the model determined from the comparison of tide gauge observations with modelled sea-surface heights was used.

4 Results

Figure 6 shows the reduced time series of tide gauge observations exemplarily at two stations. The reduction of the sea-level variation, both due to internal and external effects, is clearly visible. This is also demonstrated by the decreased standard deviation of the reduced time series in comparison with the original time series of h_{tg} (Table 2).

The reduced time series can serve as a quality check for the observations. Some spurious values that for instance may be a hint for observation errors appear as larger variations or as outliers.

From the reduced time series the secular trend can be estimated more reliable (Table 3, center part). But, since the tide gauge observations together with corresponding modelled sea-surface heights itself formed the input for the model of the external sea-level variation signal, there is no gain in information about the absolute values of T_{rel} by this technique.

Table 2: Effect of combining the tide gauge observations and the oceanographic model: Standard deviation of observation and reduced time series

Station	Standard deviation (mm)		
	h_{tg}	$h_{tg} - var_I$	$h_{tg} - var_I - var_E$
Kemi	210	87	31
Furuogr.	205	89	29
Ratan	203	90	29
Helsinki	199	91	12
Landsort	171	91	15
Warnem.	95	76	39

The analysis of tide gauge observations and sea-surface heights provided by the oceanographic model reveals small locally different trends of the sea-level heights in the order of up to 1 mm/a during the model period. A combination of the tide gauge observations with the oceanographic model removes these internal trends, and the trend differences between tide gauge stations are in general slightly closer to the long term values (Table 3, right hand part).

It is also notable that data gaps in short time series can be treated adequately by this technique,

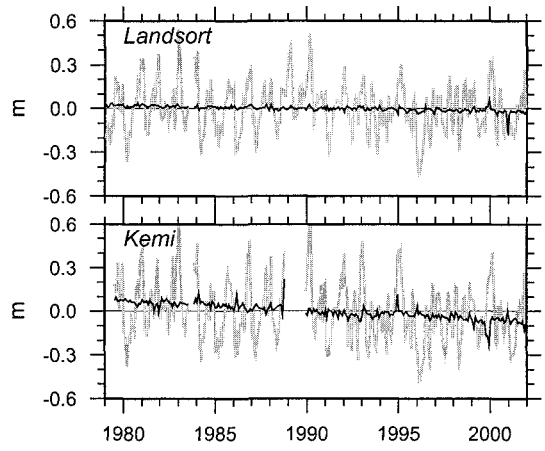


Fig. 6: Original (gray) and reduced (black) time series of tide gauge observations at two stations

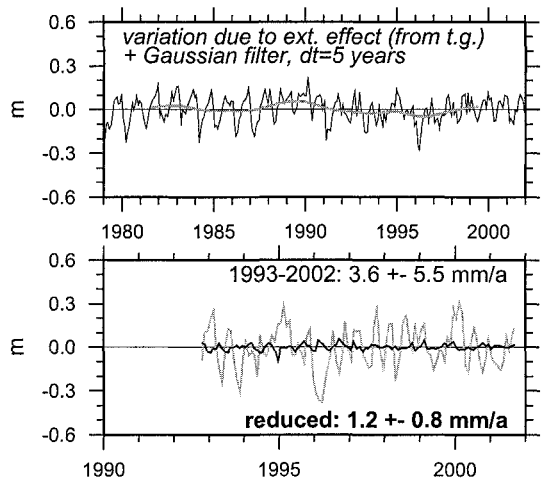


Fig. 7: Original and reduced monthly altimetric sea-surface heights
Top: Model of the external signal, that was estimated from the comparison of tide gauge observations and the oceanographic model.
Bottom: For the available altimeter data period, the absolute sea-level change rate was estimated from the original (gray) and reduced (black) data.

and a reliable trend can be estimated from these time series, too (see station Kemi in Figure 6 and Table 3).

The combination of observed and modelled sea-level heights is particularly useful for the determination of the absolute (geocentric) sea-level change from short altimeter time series. Figure 7 shows the result for the TOPEX altimetric data. From the reduced time series of TOPEX sea-surface heights the regional sea-level trend can be estimated more reliably. The figure again shows the modelled external effect in the Baltic Sea sea-level variation, which was obtained from tide gauge observations and the oceanographic model, with a low pass filter applied. The plot reveals an apparent slight positive sea-level trend within the last third of the considered time period, which leads to an overestimation of the geocentric sea-level change rate from the TOPEX altimetry data. Applying the modelled external signal this effect is also reduced in the TOPEX data.

The trend value determined from the original altimetric sea-surface heights gives 3.6 mm/a with a statistical error ($1\,\sigma$) more than the value itself. After utilising the additional information from the oceanographic model and from tide gauges, a regional sea-level rise of 1.2 mm/a is estimated, which is a more plausible value also in comparison with long-term tide gauge analyses.

5 Conclusion and outlook

Sea-surface heights in the Baltic Sea obtained from tide gauge observations, satellite altimetry and from a precise, high-resolution oceanographic model were combined in order to reduce the variation in the observation time series. While the internal effects in the Baltic Sea sea-level variations are well represented in the oceanographic model and can therefore be subtracted directly from the observations, the external effect was modelled by comparison of tide gauge observations and corresponding modelled heights.

The reduced time series of sea-surface height observations allow a quality check of the observation data. Moreover, by reducing internal effects in the Baltic Sea sea-level variation the estimation of trend differences between tide gauge stations from short time periods can be improved. Also, reliable trends can be determined from short time series affected by data gaps.

After the combination of satellite altimetry, the oceanographic model and tide gauge observations, the absolute (geocentric) regional sea-level rise for the Baltic Sea was estimated to 1.2 mm/a.

The external signal in the Baltic Sea sea-level variation was modelled with the assumption, that it has no trend in the considered time period. However, longer observations that extend the time interval of the oceanographic model are

Table 3: Trends estimated from tide gauge observations for different time periods and from the reduced time series, and trend differences relative to station Warnemünde

	Trend (mm/a) determined from			Trend difference (mm/a) relative to Warnemünde		
Station	(1)	(2)	(3)	(1)	(2)	(3)
Kemi	-6.8±0.3	-5.8±2.0	-6.7±0.3	-8.1	-7.0	-7.8
Furuogrund	-8.0±0.3	-6.9±1.9	-8.1±0.3	-9.3	-8.1	-9.2
Ratan	-7.6±0.3	-7.3±1.8	-8.1±0.2	-8.9	-8.5	-9.2
Helsinki	-1.4±0.3	-1.6±1.8	-1.4±0.1	-2.7	-2.8	-2.5
Landsort	-2.5±0.3	-2.4±1.6	-1.9±0.1	-3.8	-3.6	-3.0
Warnemünde	1.3±0.2	1.2±0.9	1.1±0.4			

(1) ... tide gauge observations, 1930-2002
(2) ... tide gauge observations, 1979-2002
(3) ... reduced time series, 1979-2002

necessary to relate the modelled external signal to the long-term sea-level behaviour.

The study showed that the combination of the different sea-level information can make use of the strength of each data source to reduce the variation in the sea-surface height time series and to separate one phenomenon of interest. The technique described may also be useful for a number of other geodetic applications, such as altimeter validation.

Acknowledgment

This project was partly funded by the German Federal Ministry of Education and Research (BMBF) (FKZ: 01LD0025).

References

Dietrich, R. and Liebsch, G. (2000). Zur Variabilität des Meeresspiegels an der Küste von Mecklenburg–Vorpommern. *Zeitschrift für geologische Wissenschaften*, 28(6):615–623.

Ekman, M. (1988). The world's longest continued series of sea level observations. *Pure and Applied Geophysics*, 127:73–77.

Ekman, M. (1992). Postglacial rebound and sea level phenomena, with special reference to Fennoscandia and the Baltic Sea. In *Nordic Geodetic Commission, Autumn School "Geodesy and Geophysics", Helsinki,Finnland, September 7-13, 1992.*

Ekman, M. (1996). A common pattern for interannual and periodical sea-level variations in the Baltic Sea and adjacent waters. *Geophysica, Finnish Geophysical Society, Helsinki*, 32(3):261–272.

Ekman, M. (1998). Secular change of the sea level variation in the Baltic Sea and secular change of the winter climate. *Geophysica*, 34(3):131–140.

Forsberg, R., Kaminskis, J., and Solheim, D. (1996). Geoid of the Nordic and Baltic region from gravimetry and satellite altimetry. In Segawa, J., Fujimoto, H., and Okubo, S., editors, *IAG Symposia On Gravity, Geoid and Marine Geodesy (GraGeoMar) 1996, Tokyo, Japan, Sep. 30 - Oct. 5., 1996*, volume 117 of *Springer series of IAG Symposia*, pages 540–547. Springer.

Lehmann, A. (1995). A three-dimensional baroclinic eddy-resolving model of the Baltic Sea. *Tellus*, 47A:1013–1031.

Lehmann, A. and Hinrichsen, H.-H. (2000). On the thermohaline variability of the Baltic Sea. *J. Mar. Sys.*, 25:333–357.

Liebsch, G. (1997). *Aufbereitung und Nutzung von Pegelmessungen für geodätische und geodynamische Zielstellungen.* Deutsche Geodätische Komission, Reihe C, Nr. 485, 108 Seiten.

Novotny, K., Liebsch, G., Dietrich, R., and Lehmann, A. (2002a). Sea-level variations in the Baltic Sea: Consistency of geodetic observations and oceanographic models. In Ádám, J. and Schwarz, K.-P., editors, *Vistas for Geodesy in the New Millenium. IAG Scientific Assembly, Budapest, Hungary, September 2–7, 2001*, volume 125 of *Springer Series of IAG Symposia*, pages 493–498. Springer.

Novotny, K., Liebsch, G., Dietrich, R., and Lehmann, A. (2002b). Studying seasonal sea-level variations in the Baltic Sea using geodetic measurements and an oceanographic model. 50 years Geodetic Science at The Ohio State University, Weikko A. Heiskanen Symposium, Columbus, Ohio (USA), Oct. 1-4, 2002.

Samuelsson, M. and Stigebrandt, A. (1996). Main characteristics of the long-term sea level variability in the Baltic Sea. *Tellus*, 48A(5):672–683.

Analysis of GPS Campaign Observations of the Asia Pacific Regional Geodetic Project (APRGP) -Establishment of a Velocity Field-

YOSHIYUKI TANAKA, TOMOMI AMAGAI, AKIKO YAMADA, SHIGERU MATSUZAKA, AND HIROMICHI TSUJI

Geographical Survey Institute, 1, Kitasata, Tsukuba, Japan

ytanaka@gsi.go.jp, 81-29-864-4804, 81-29-864-1802

Abstract. Cooperative campaigns by space geodetic techniques have been actively carried out around the world to develop global GIS infrastructures. Asia Pacific Regional Geodetic Project (APRGP), as a regional campaign, has promoted annual observations mainly by GPS to densify the International Terrestrial Reference Frame (ITRF). To construct an ITRF based frame, not only positions but also velocities are necessary. Nevertheless, establishing a velocity field in the region has not been achived yet. In this paper, we estimate a velocity field, using GPS data obtained from 1998 to 2001 campaings. The result shows an excellent agreement with the plate model.

Keywords: Asia Pacific Regional Geodetic Project, APRGP, ITRF, PCGIAP

1 Introduction

The Permanent Committee on GIS infrastructure for Asia and the Pacific (PCGIAP) is a forum of national mapping organizations in the region to maximize the economic, social and environmental benefits of geographic information. Since unified positional information plays a key role in GIS developments, PCGIAP has a working group on Regional Geodesy to enhance a regional geodetic infrastructure through annual cooperative campaigns. The working group headed by Geoscience Australia has organized the Asia Pacific Regional Geodetic Project (APRGP) campaigns shaping in 1997 in collaboration with member nations in the region (PCGIAP working group 1, 1998). As an active member of the working group, the Geographical Survey Institute Japan (GSI) joined the campaigns and processed the data to yield sets of coordinates of occupied sites. These are all submitted to Geoscience Australia for the densification of the International Terrestrial Reference Frame (ITRF). In a similar way, various countries have submitted the data and the site coordinates obtained in each annual campaign (PCGIAP working group 1, 1998, 1999).

To construct an ITRF based frame, however, not only positions but also velocities are necessary. Moreover, the science of the complex plate tectonic behavior as well as disaster prevention will benefit from a precise knowledge of the velocity field. In this paper, we will report the determination of a velocity field in the region using GPS data.

2 Data and analysis method

GPS data acquired in four campaigns from 1998 to 2001 are used in the analysis. In each campaign, about one hundred sites have been occupied for about ten days. One hundred and thirty sites were observed more than once, for which velocities were available. About ten different antenna types were used in the network and choke ring antenna types were evenly distributed in the entire region, representing 34 % of the total number of sites.

The analysis method is as follows. Our goal was to obtain a solution satisfying two conditions. First, it should form a precise network in the region. Second, it should be consistently embedded onto the global reference frame. We chose GAMIT&GLOBK software for this task (King and Bock, 2000).

To avoid unnecessary systematic error, the analysis strategy was consistent through the campaigns. The first key point of the strategy was that sites were divided into several clusters because

the number of campaign sites was too large. Clustering was performed considering antenna types, proximity of stations and existence of IGS sites near the campaign ones. The numbers of clusters (ten on the average) were slightly different through the campaigns, but the way of clustering was basically unchanged. GAMIT computation (version 10.07 for the 1998 and 2000 campaigns, 10.04 for the 1999 and 2001 campaigns) was then performed for each cluster to obtain a daily biases-fixed solution with loose constraints (King and Bock, 2000). IGS final orbits were used and the elevation cut-off angle was 15 degree. Tropospheric delays were estimated every two hours with Niell's mapping functions for the hydrostatic and water-vapor delays and the gradient parameters were not estimated.

The second point is that the reference frame was defined using a biases-free solution from SIO (Scripps Institution of Oceanography) global analyses. These daily global solutions are combined with those daily solutions of the campaign sites by GLOBK 5.06I into a file containing annual positions and covariance matrix information. Finally, the solutions for four campaigns are combined and the velocity field was determined. In this process, the orientation, translation and scale of the network as well as their time derivatives were adjusted. The orbital parameters and the Earth rotation parameters were also estimated. Site coordinates and velocities were loosely constrained with their vertical tolerances loosened compared to the horizontal ones.

We adopted ITRF97 (Boucher et al., 1997) as the reference frame to compare the velocities with the recent plate model in the next section (This is also because our previous analyses of APRGP98-00 were done with ITRF97 and a recalculation with ITRF2000 (Boucher et al., 2003) takes considerable time.). To materialize the reference frame, the sites whose coordinates and velocities were precisely estimated (Dong et al., 2002) were used to fix the origin of the network. As a result, all parameters (including the coordinates and velocities of IGS sites) were adjusted such that the deviations from a priori values were minimized considering the a priori standard deviations.

3 Results

Figure 1 shows the resultant horizontal velociy field in the entire region. Because the number of the stations is too large, some sites close to each other are omitted. The black and white circles represent IGS stations and APRGP sites, respectively. Error ellipses show the interval at a confidence level of 95 %. It is seen that the velocities of campaign sites are analogous to those of their neighboring IGS stations (The vertical ones are not shown because the velocity values at most sites are much smaller than the corresponding standard deviations.).

Next, let us magnify the velocity field of Southeast Asia, where the plate motion is comparatively active (Figure 2). Several sites are removed to make the figure easily viewable. The curve shows a plate boundary. Superimposed white arrows represent the recent plate velocity model 'REVEL' (Sella et al., 2002), which was mainly derived from eight years of GPS observations (from 1993 to 2000). The plate model and the velocity field agree very well with each other, both in direction and magnitude. It is well understood that the campaign sites have an important role to fill the gap between IGS stations.

Finally, the mean standard deviations are shown in Table 1. Standard deviation values for most of the APRGP sites are at the level as several millimeters per year. On the other hand, those for IGS sites used to define the reference frame are about two times as large as their a priori values. This means that the inconsistency between the global reference frame and the regional network is a few millimeters per year. This inconsistency might be caused by two reasons: less accuracy of campaign sites and a large number of campaign sites compared to IGS sites. To improve the standard deviations, effects of clustering may have to be studied. Because this takes a considerable computation time, we didn't treat in this paper.

4 Conclusion

A densified velocity field in Asia and the Pacific region from GPS data was obtained. The velocity filed shows an excellent agreement with both that of the neighboring IGS stations and that of the plate model derived from recent GPS data. This means that the campaign sites can play an important role to supplement IGS stations and densify ITRF in the region. In the future, we would like to improve the velocity field combining the 1997 and 2002 campaigns data and other space geodetic techniques such as VLBI, SLR, etc in ITRF2000.

Acknowledgements. We are grateful to Geoscience Australia.

References

Boucher, C., Z. Altamimi, and P. Sillard (Eds.) (1997) The 1997 International Terrestrial Reference Frame (ITRF97), IERS Tech. Note 27, Obs. De Paris.

Boucher, C., Z. Altamimi, P. Sillard and Martine Feissel-Vernier (2003) The ITRF2000, IERS Tech. Note 31, Obs. De Paris.

Dong, D., P. Fang, Y. Bock (2002) M. K. Cheng, and S. Miyazaki, Anatomy of apparent seasonal variations from GPS-derived site position time series, JGR, 107, No. B4, 10.1029/2001JB000573, 2002, ETG 11, 1-12

King, R. W., and Y. Bock (2000) Documentation for the GAMIT GPS analysis software, Mass. Inst. of Technol., Cambridge, Mass.

Sella, F. G., T., Dixon, and A. Mao (2002) REVEL: A model for Recent plate velocities from space geodesy, JGR, 107, No. B4, 10.1029/2000JB000033, 2002, ETG 9, 1-17

Working group 1 of the PCGIAP (1998) Proceedings of the workshop on regional geodetic network, Australia, July, 1998.

Working group 1 of the PCGIAP (1999) Proceedings of the second workshop on regional geodetic network, Vietnam, July, 1999.

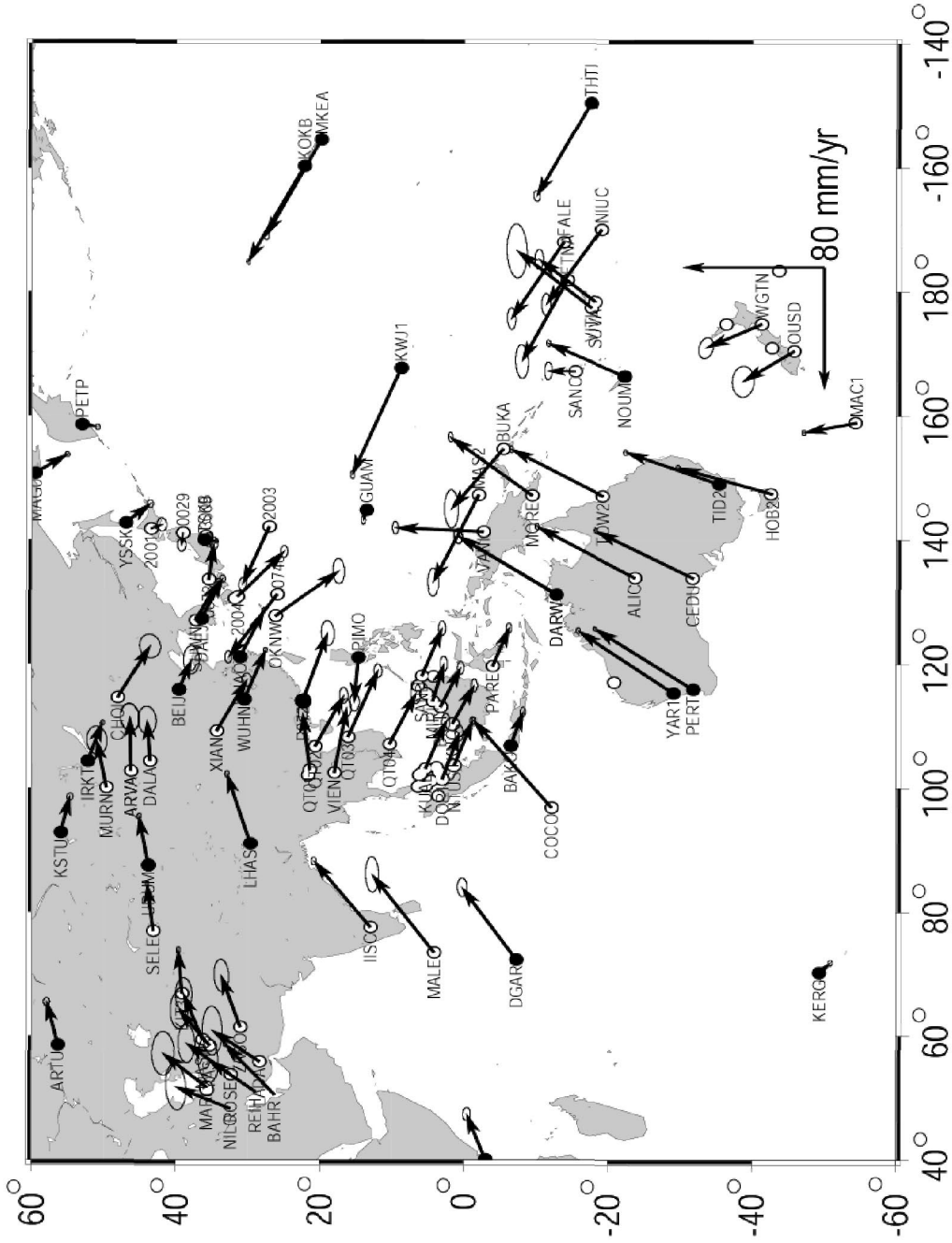


Fig. 1 The horizontal velocity field. The black and white circles show IGS and the campaign sites, respectively. The result for sites occupied in at least two campaigns is shown.

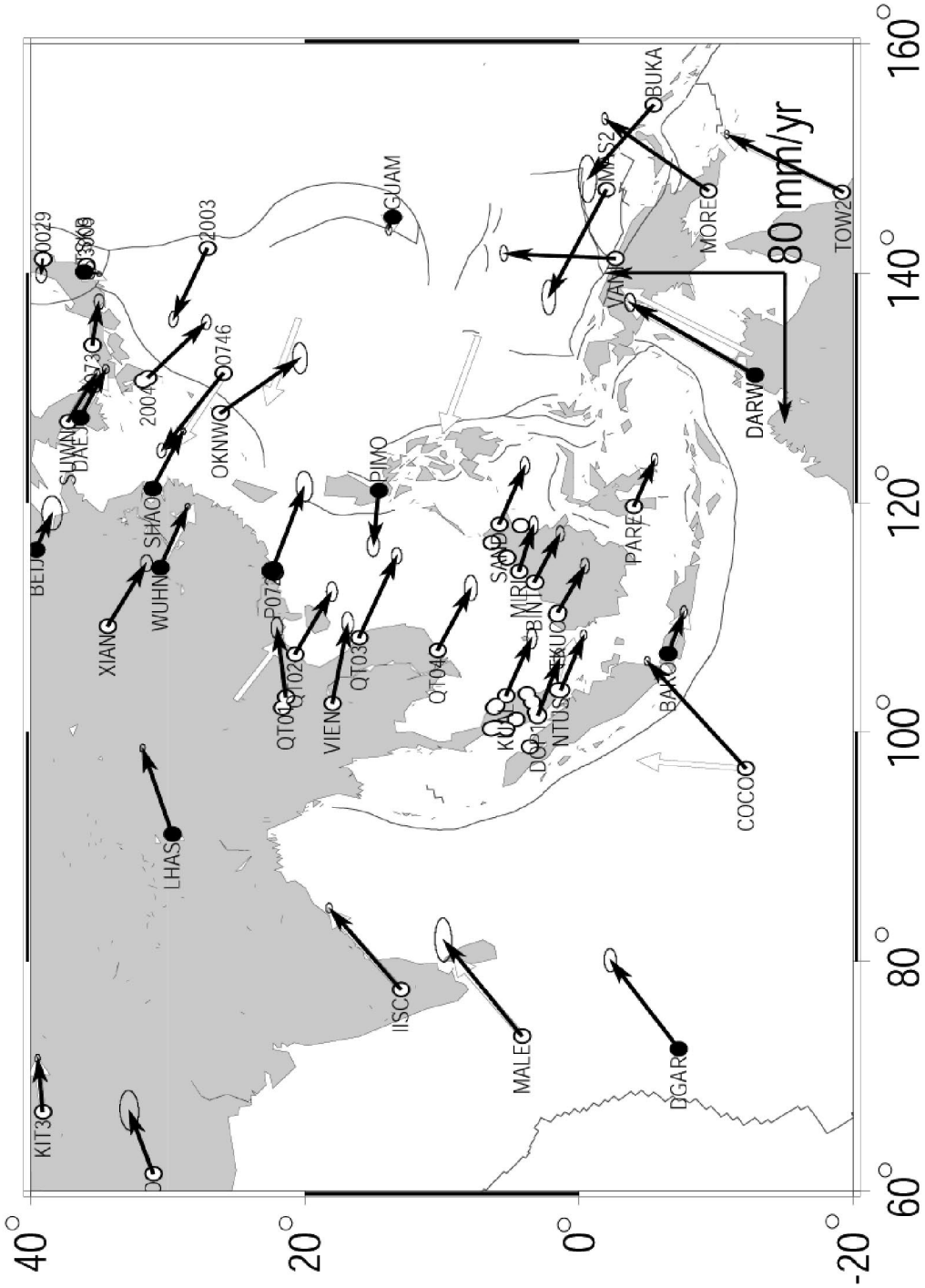


Fig. 2 The magnified version of Figure 1. The white arrow shows the plate velocity model (Sella, et al., 2002).

Table 1. Mean standard deviations of the velocity components (mm/year)

Sites	Sigma E	Sigma N	Sigma H
APRGP	8.4	5.4	12.9
IGS (except below)	4.8	2.1	6.3
IGS (used to fix the origin)	3.1	1.8	6.5

The proposed International DORIS Service

Pascal Willis
Institut Géographique National, Direction Technique, 2, Avenue Pasteur, BP 68, 94160 Saint-Mandé, France
Jet Propulsion Laboratory, MS 238-600, 4800 Oak Grove Dr, Pasadena CA 91109, USA

Gilles Tavernier
Centre National d'Etudes Spatiales, 18, Avenue E. Belin, 31401 Toulouse Cedex 9, France

Martine Feissel-Vernier
Observatoire de Paris, 61, Avenue de l'Observatoire, 75014 Paris, France
Institut Géographique National, LAREG, 6-8, Avenue Blaise Pascal, 77455 Marne-la-Vallée, France

Frank Lemoine, Carey Noll
NASA Goddard Space Flight Center Code 920, Laboratory for Terrestrial Physics, Greenbelt MD 20771, USA

John Ries
CSR, Mail Code R1000, The Univ. of Texas at Austin, Austin TX 78712, USA

Laurent Soudarin
CLS, 8-10 rue Hermes, 31526 Ramonville Cedex, France

Abstract In 1999, an International DORIS Experiment was initiated under the aegis of the International Association of Geodesy. Since then, 3 new DORIS satellites have been launched carrying improved Doppler receivers. In addition, the DORIS tracking network has also been steadily improved. In 2001, a DORIS analysis campaign was launched. Several groups participated using different software and providing weekly and monthly time series of results, making them available on Internet. The goal of this paper is to discuss the status of the DORIS Pilot experiment, to analyze the progress made during the past 2 years and to propose the official creation within the IAG of an International DORIS Service (IDS). Such a service would operate in conjunction with the already existing IGS, IVS and ILRS as part of the proposed International Global Geodetic Observing System (IGGOS).

Keywords. DORIS, IDS, International DORIS Service, IGGOS

1 Introduction

Since the launch of the first DORIS receiver on-board the SPOT-2 satellite in 1990, substantial progress has been made in equipment technology as well as

in software improvements. A large international community now uses DORIS data for different types of scientific activities: precise orbit determination (Tapley et al., 1994, Davis, et al. 1997, Haines et al., 2003), geodesy, geodynamics (Crétaux et al., 1998, Soudarin et al., 1999, Crétaux et al., 2002, Willis et al., 2003b), realization of Terrestrial Reference Frame (Altamimi et al., 2002), Earth rotation, ionosphere, space weather, temporal gravity field. By comparison with other techniques, the DORIS technique has several advantages, thanks to the low altitude of its satellite (geocenter determination, lower ionospheric measurements) and to the density of its tracking network (separability of long wavelengths in gravity fields solutions). In 1999, a DORIS Pilot Experiment (Tavernier et al., 2002) was initiated within the IAG to prepare a possible additional Service of the International Association of Geodesy, based on the same principles as the IGS and aiming for similar scientific applications (Beutler et al., 1999) and chaired by G. Tavernier. These IAG services (IGS, IVS, ILRS and IDS), in conjunction with IERS would be the backbone of a future International Geodetic Observing System, as proposed by Rummel (2002) and Beutler (2002). The goal of this paper is to present the recent developments in the DORIS technology, to summarize the results and the experience gained from the DORIS Pilot Experiment and to propose a new International DORIS Service.

1.1 DORIS satellites and receivers

During the past two years, three more DORIS satellites were launched: Jason (CNES/JPL for oceanography) on December 7, 2001, ENVISAT (ESA for oceanography) on March 1, 2002 and SPOT-5 (CNES for remote sensing) on May 4, 2002. Today there are six available DORIS satellites. This doubling of the available DORIS satellites has enhanced the positioning performance of the DORIS system (Willis et al., 2003b). Furthermore, these new DORIS receivers have a lower Signal-to-Noise Ratio, and they are equipped with a dual channel capability allowing common observations of two DORIS beacons within the same geographic region (Tavernier et al., 2003). The improved on-board software DIODE (Jayles et al., 2002) allows precise real-time positioning of the satellites, currently at the sub-meter level in 3-D in real-time. Following the experience gained with the new satellites, some recent software updates will be made to all new DORIS satellites before the end the year, as already done for SPOT-4. It must also be noted that excellent orbit results were obtained for the TOPEX and Jason satellites when combining DORIS data with SLR and/or GPS data (Haines, et al., 2003; Luthcke et al., 2003).

In the near future, additional satellites will carry these new types of DORIS receivers: Cryosat (ice sheet altimetry mission, 2004), Jason-2 (oceanography mission, 2005) and potentially several satellites of the Pleiades constellation (SPOT follow-on remote sensing mission, 2007).

1.2 DORIS network and data centers

During the past 10 years, the DORIS permanent tracking network has been extremely stable, making this system attractive for long-term geodetic and

geophysical studies such as reference frame monitoring (Altamimi et al., 2002). The network is presently composed of 56 stations geographically well distributed over 33 countries as displayed in Figure 1. Some new stations were added to further improve the already good geographical distribution of the tracking network (e.g., Sal Island, Cap Verde; Thule, Greenland). Since 2000, a very important rejuvenation of the network has been undertaken by the Institut Géographique National: recent local ties campaigns, improved monumentation, and local investigations on long-term antenna stability (Fagard, 2003). Stations, presently assessed with a poor or dubious monumentation (Figure 1) will be upgraded to more recent geodetic standard in the near future.

Fig 1 The DORIS permanent tracking network monumentation classification (July 2003) : E = excellent, G = Good, D = Dubious, P = Poor.

Presently, 26 out of the 56 DORIS stations are considered as excellent from the standpoint of these criteria. The newly installed DORIS stations can now slightly shift their transmitting frequency by +/- 50 kHz on the 2 GHz frequency and +/- 10kHz on the 400 MHz frequency in order to avoid interference with nearby stations, as viewed from the orbiting DORIS receivers. Such third generation beacons are now installed in twelve stations: Cibinong, Easter Island, Kauai, Mahe, Manilla, Marion Island, Santa-Helena, Sal, Terre Adélie, Thule, Toulouse, Tristan Da Cunha. Following some problems related to the DORIS/Jason oscillator (Willis et al., 2003b), a third DORIS Master Beacon (reference for time and frequency) will be soon installed in Hartebeesthoek (South Africa).

With the new dual-channel capability, CNES has proposed to extend this network and sent out a call for participation. A group has been formed to handle a possible extension of the present DORIS ground network, under the chairmanship of K. Larson (University of Colorado), then F. Lemoine (Goddard Space Flight

Center). For example, a first DORIS beacon has been installed at Wettzell in Germany in order to prepare a collocation with the TIGO project in South America (Schlüter et al., 2002). A specific campaign (Lambert campaign) has also been conducted on the Sorsdal glacier from December 12, 2001 to February 5, 2002, led by Geoscience Australia (Govind et al., 2002) to measure ice deformation. More recently, a new similar campaign was conducted on the Lambert glacier in January 2003. Another campaign was organized in Corsica by the CERGA for the Jason/TOPEX calibration activities. These additional DORIS data, obtained as a densification of the permanent tracking network, may soon be distributed for further investigations. In the future, a similar campaign will also be conducted at the Gavdos calibration site in Crete in collocation with SLR, GPS and tide gauge instruments (Pavlis et al., 2003).

The DORIS data distribution has been improved in the past year. The CNES data center SSALTO has been able to cope with the recent doubling in the DORIS data stream. The IGN data center is now operational as a mirror to the CDDIS data center that has been performing well since 1993. Those two data centers provide preprocessed DORIS data from all satellites on a daily basis, as well as necessary information files (satellite maneuvers, attitude files, physical models, etc). More recently, DORIS results (weekly and monthly stations positions, long-term stations positions and velocities, geocenter variations, ionospheric contents) from different analysis centers can also be obtained at these data centers. The DORIS Analysis Coordinator (M. Feissel-Vernier) in conjunction with the DORIS Data delivery Coordinator (C. Noll) conducted a study to better define the format, content and delivery specifications.

2 From the DORIS Pilot experiment to the IDS

The DORIS Pilot Committee is steered by an international committee, chaired by G. Tavernier. The day-to-day activities, such as the Web site maintenance (http://ids.cls.fr), are conducted by a Central Bureau (CLS, CNES, and IGN). A DORISMail facility has been created and is presently distributing mail to 158 individuals. At the end of June 2003, a total of 246 emails were distributed and are still archived and easily accessible at http://list.ensg.ign.fr/wws/arc/dorismail. Anybody can subscribe or access those previous messages.

Several international meetings were organized recently. All presentations can be found at the IDS Central Bureau Web page (http://ids.cls.fr/html/report.html): e.g. IDS Workshop, Biarritz, June 13-14, 2002 (90 participants), and IDS Analysis workshop, Marne-la-Vallée, February 20-21, 2003 (30 participants). The purpose of these workshops is to exchange experience between analysis groups to improve results from all groups. The Analysis Coordination website (http://lareg.ensg.ign.fr/IDS) also gives access to IDS information, data and results.

As the DORIS results are now available regularly on the Internet, several groups have used them in conjunction with results from other techniques (e.g., Nothnagel et al., 2002). Of course additional steps are needed to increase the accuracy of the results as well as their integrity, for example by first

unifying all DORIS solutions, but it must be noted that a somehow important activity has been conducted in the past two years by the DORIS Analysis groups. Furthermore, some major improvements are anticipated when new gravity field models, based on recent gravity field missions, such as GRACE (Tapley et al., 2002) are available to all participants. The accuracy of these solutions should enhance the accuracy of the DORIS geodetic results as Low Earth Orbit satellites such as SPOT at 830 km altitude and more recently ENVISAT at 750 km altitude.
A proposal for an International DORIS Service was presented to the IAG in Sapporo, for an organization that would be very similar in principle with the already well-functioning IGS (Beutler et al., 1999) or the more recently created IVS and ILRS.

3 Status of the IDS Analysis campaign

On November 2001, an international call for participation was issued by the DORIS Pilot Experiment Central Bureau to process the DORIS measurements for precise point positioning (time series of coordinates). Six out of the eleven DORIS analysis centers answered positively to process the data (Table 1).

Table 1. Participating IDS Analysis Centers (as in July 2003)

Analysis Center Software	Contact	Orbits	Stations coord.	Geocenter	EOP
CNES ZOOM	J.P. Berthias	X	X		
CSR UTOPIA	J. Ries	X	X		
IAA	E. Yagudina				
IGN/JPL GIPSY-OASIS	P. Willis	X	X	X	X
INASAN GIPSY-OASIS	S. Tatevian	X	X		
LEGOS/CLS GINS	J.F. Crétaux	X	X		X
SSALTO	G. Tavernier	X	X		
ESA/ESOC	J. Dow				
Geodetic Obs. Pecny	J. Kostelecky				
Geoscience Australia MICROCOSM	R. Govind	X	X		
Royal Obs. of Belgium	R. Warnant				
Univ. Of Newcastle	P. Moore				

Results were presented at the IDS Analysis workshop using different software at the 1 to 2 cm level for global solutions. Experiences were exchanged during the meetings and a preliminary

combination of all available series was obtained for weekly and monthly solutions. Several problems inherent to the DORIS techniques and to some specific DORIS stations were raised (Earthquakes, antennas falls). Data formats and distributions were discussed leading to new standards. These discussions should help us provide a regular unified DORIS solution obtained by a rigorous combination of all available DORIS individual time series solutions.

4 Conclusions

In conclusion, the DORIS technique has benefited in the past two years from significant improvement in the receiver and transmitter technology as well as from a large increase in the number of available satellites. This has attracted new groups to processing the DORIS data for geodetic or geophysical purposes. This community joined its forces within the scope of the IAG established DORIS Pilot Experiment in 1999. Since then, several international meetings have been organized and regular results from different analysis groups are regularly available on Internet. We think that the time has now come to create a new IAG Service, to be called International DORIS Service, which would be naturally a part of the soon-to-come Integrated Global Geodetic Observing System (IGGOS).

Acknowledgements

Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

Altamimi Z, Sillard P, Boucher C (2002) ITRF2000, a new release of the International Earth Rotation System. J. Geophys. Res. Solid Earth 107: B10 2214.

Beutler G, Rothacher M, Schaer S, Springer TA, Kouba J, Neilan RE (1999) The International GPS Service (IGS, An interdisciplinary service in support of Earth Sciences). Adv. Space Res. 23: 4 631-653.

Beutler G, Brunner F, Dickey J, Feissel M, Forsberg R, Mueller II, Rummel R, Sanso F, Schwartz KP (2002) The IAG review 2000-2001: Executive Summary. International Association of Geodesy Symposia, Springer-Verlag 125: 603-608.

Crétaux JF, Soudarin L, Cazenave A, Bouille F (1998) Present day tectonic plate motions and crustal deformations from the DORIS space system. J. Geophys. Res. 103: 30167-30181.

Crétaux JF, Soudarin L, Davidson FJM, Gennero MC, Bergé-Nguyen M, Cazenave A (2002) Seasonal and interannual geocenter motion from SLR and DORIS measurements: comparison with surface loading data. J. Geophys. Res. 107.

Davis GW, Ries JC, Tapley BD (1997) The accuracy of precise orbits computed from Doppler tracking data. J. Astr. Sci. 45 : 4 451-469.

Fagard H (2003) The DORIS network, status report on renovation and collocations. IDS Analysis Workshop, IGN, Marne-la-Vallée, France, February 2003.

Govind R, Valette JJ, Dawson J (2002) The Antarctic Sordal Glacier Experiment. IDS Workshop, Biarritz, France, June 2002.

Haines BJ, Bertiger WIB, Desai S, Kuang D, Munson T, Young L, Willis P (2003) Initial Orbit Determination Results for Jason-1: Towards a 1-cm Orbit J. Navigation, 50:3 171-179.

Jayles C, Berthias JP, Laurichesse D, Nordine S, Cauquil P, Tavernier G (2002) DORIS-DIODE: Two-years results of the first European Navigator. Adv Space Res 30: 2 301-306.

Luthcke, SB, Zelensky, NP Rowlands DD, Lemoine FG, and Williams TA (2003), The 1-cm orbit: Jason-1 precision orbit determination using GPS, SLR, DORIS and altimeter data, Marine Geod, 26 :3-4, in press.

Nothnagel A, Dill R, Feissel-Vernier M, Ferland R, Noomen R, Willis P, (2002) EOP Alignment Campaign, IDS/IGS/ILRS/IVS EOP combinations and systematic errors, IERS Workshop, Munich, Novembre 2002.

Pavlis E.C., Mertikas SP, Evans K (2003) Sea level, tectonics, environmental monitoring and.altimeter calibration in eastern Mediterranean, EOS, Trans. AGU 84(46), Fall Meet. Suppl., Abstract G12C-0284.

Rummel R, Drewes H, Beutler G (2002) Integrated Global Geodetic Observing System (IGGOS): A candidate IAG project. International Association of Geodesy Symposia, Springer-Verlag 125: 609-614.

Schlüter, W, Hase H, Böer A, Riepl S, Cecioni A (2002) TIGO starts its operation in Concepcion in 2002, CSTG Bulletin 17: 92-93.

Soudarin L, Crétaux JF, Cazenave A (1999) Vertical crustal motions from the DORIS space geodesy system., Geophys. Res. Lett. 26: 1207-1210.

Tapley BD, Ries JC, Davis GW, Eanes RJ, Schutz BE, Shum CK, Watkins MM, Marshall JA, Nerem RS, Putney BH, Klosko SM, Luthcke SB, Pavlis D, Williamson RG, and Zelensky NP (1994) Precision Orbit Determination for TOPEX/POSEIDON, J Geophys Res, 99 (C12) 24383-24404.

Tapley BD (2002) The GRACE mission: Status and performance assessment, EOS Trans., AGU, 83:47, Fall AGU Meeting Suppl, Abstract G12B-01, 2002.

Tavernier G, Soudarin L, Larson K, Noll C, Ries J, Willis P (2002) Current status of the DORIS Pilot Experiment. Adv. Space Res. 30: 2 151-156.

Tavernier G, Granier JP, Jayles C, Sengenes P, Rozo F (2003) The Current Evolutions of the DORIS System. Adv. Space Res., 31: 8 1947-1952.

Willis P, Haines B, Bar-Sever Y, Bertiger W, Muellerschoen R, Kuang D, Desai S (2003a) Topex/Jason combined GPS/DORIS orbit determination in the tandem phase. Adv. Space Res., 31:8 1941-1946.

Willis P, Bar-Sever Y, Tavernier G (2003b) DORIS as a potential part of a Global Observing System. Springer-Verlag, Proc. IUGG General Assembly, Sapporo, July 2003, submitted to J. Geodyn.

Calibration/validation of GOCE data by terrestrial torsion balance observations

Gy. Tóth[1], J. Ádám[1], L. Földváry [1,4], I.N. Tziavos[2], H. Denker[3]

[1] *Department of Geodesy and Surveying, Budapest University of Technology and Economics, P.O.B. 61, H-1521 Budapest, Hungary.* and *Physical Geodesy and Geodynamics Research Group of BUTE-HAS*
[2] *Department of Geodesy and Surveying, Aristotle University of Thessaloniki, Univ. Box 440, 54124 Thessaloniki, Greece*
[3] *Institute für Erdmessung, Universität Hannover, Schneiderberg 50, D-30167 Hannover, Germany*
[4] *Institute for Astronomy and Physical Geodesy, Technical University of Munich, D-80290 Germany*

Abstract One promising method for the external validation and calibration of the upcoming GOCE satellite mission data is the use of ground gravity field data continued upward to satellite altitude. There is a unique situation for Hungary in this respect since surface gravity gradients are available at 20143 points over an approximately 48700 km^2 area, measured by the classical Eötvös torsion balance. The concept of this contribution is to test the usability of these point gravity gradient observations for upward continuation to the GOCE satellite orbit in combination with different geopotential models and other gravity field information.

The computations are based on the least squares collocation method and the direct numerical integration of the torsion balance data. For the latter method, the spectral combination technique and the classical integration kernels are considered. Furthermore, various other data sources, such as the T_{zz} gravity gradients based on the gravity and terrain data collected within the frame of the European Geoid Project, are utilized for comparisons.

Besides the comparisons between the different satellite gravity gradient computations, an error analysis of the results is presented.

Keywords. GOCE, vertical gravity gradients, upward continuation, Eötvös' torsion balance

1 Introduction

Ground gravity field data will be useful for calibration/validation of the gravity gradients that are to be measured by the upcoming GOCE satellite gravity field mission (Arabelos and Tscherning, 1998; Denker, 2002). In this paper the possibility of using ground gravity field gradients for calibration of Level 1b gravity gradients of GOCE is investigated by using gravity field information provided by the Eötvös torsion balance. The intent is instrument calibration and/or data validation.

As it is well known, torsion balance data provide short wavelength gravity field information, and therefore it is interesting whether these data can be used in the context of space gravity field data calibration. It is because short wavelength information is mainly lost when we upward continue our ground gravity field data. Another question is whether the spatial extension of our data is enough for upward continuation to the GOCE altitude (approximately 250 km). In this respect the truncation error is expected to be quite critical. The issue of the combination of gravity anomalies and gravity gradient data is also important to be investigated, and the question is whether this technique can yield accurate enough gravity field predictions for the GOCE orbit.

The algorithms used in this study were the least squares collocation and spectral combination method. We have used collocation because of its flexibility, particularly its capability to use and to predict various gravity field quantities and since it provides statistical quantities, estimates of covariances and cross-covariances

2 Torsion balance data

The original design of the Eötvös torsion balance goes back to 1890. Since then a great number of such measurements were carried out in Hungary, mainly for oil field exploration. The approximate number of the measurement sites is 60000 in Hungary, but only about 20000 of them are stored in a database. The processing of the old field books is still going on at the Loránd Eötvös Geophysical Institute. The data distribution can be seen in Fig. 1. In the plot also those areas are marked where there are still unprocessed field books of measurements exist. At each station the topographic effect of the nearby and remote terrain was also calculated and stored in the database. The standard error of the W_{xz} W_{yz} data is estimated to be about 1.2 E (Eötvös) and this figure is slightly greater for the topographic

effect. The $W_{yy\text{-}xx}$ W_{xy} gravity gradients have a greater error of about 2 E. Throughout this study 2 E error standard deviation was used for the spectral combination computations using the W_{xz} W_{yz} data and 1 and 2 E for the collocation computations using both W_{xz} W_{yz} and $W_{yy\text{-}xx}$ W_{xy} gravity gradients, respectively. The statistics of all the 80130 torsion balance gravity gradients and their topographic effects are shown below in Table 1. In all further computations the topographic effects were removed from the gradients and also the normal effect on the W_{xz} and $W_{yy\text{-}xx}$ observations were taken into account (see for example Hein,1981, p 21).

Table 1. Statistics of Hungarian torsion balance data and their topographic effects. All figures shown are in E (Eötvös).

Data set	no.	min	max	mean	std
W_{xz}	20128	-143.5	135.7	8.2	12.7
t_{xz}	20128	-97.8	56.5	0.1	4.6
W_{yz}	20128	-108.6	100.4	-0.6	12.1
t_{yz}	20128	-42.7	55.2	-0.3	4.6
$W_{yy\text{-}xx}$	19936	-778.5	455.8	-1.8	29.8
$t_{yy\text{-}xx}$	19936	-292.6	204.0	-0.7	17.4
$2W_{xy}$	19938	-814.3	376.4	1.3	27.7
$2t_{xy}$	19938	-263.9	715.4	-0.9	17.3

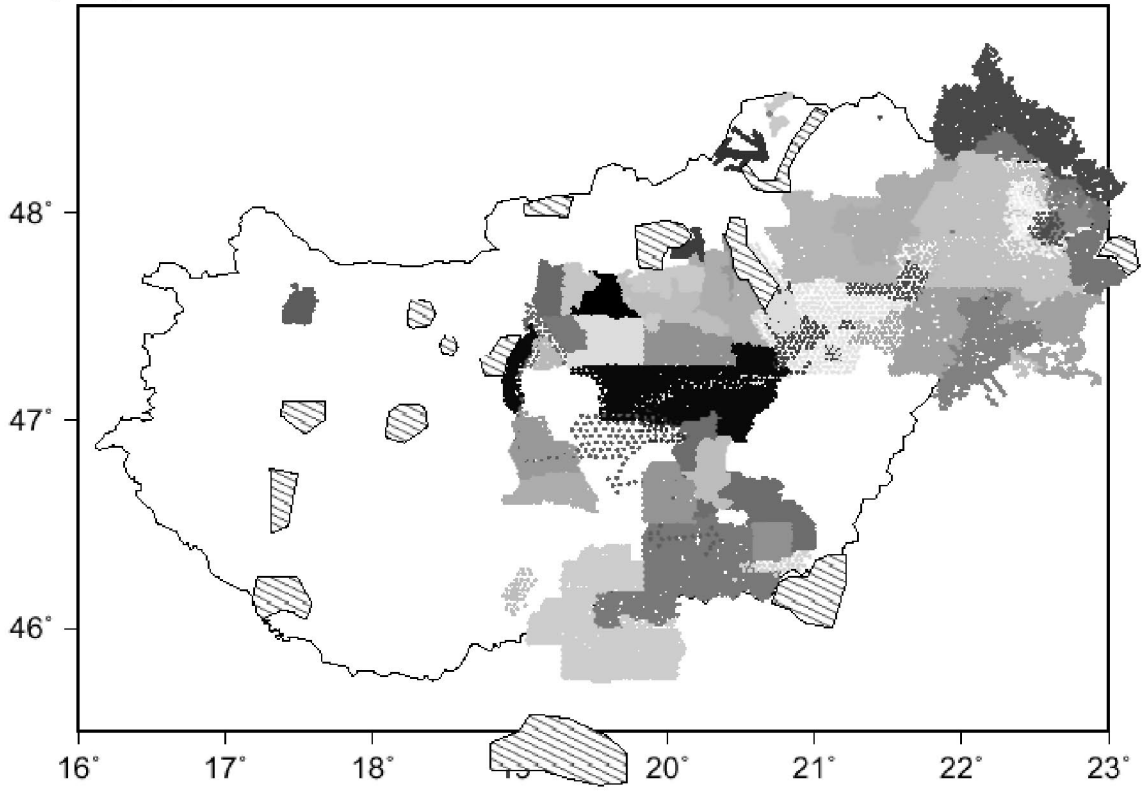


Fig. 1. Location of torsion balance measurements in Hungary (status: January 2003.). Hatched areas indicate places where torsion balance measurements exist but not yet stored in the database

3 Methodology

Least squares collocation.

The standard formulation of the well-known collocation method is not repeated here. The reader is referred to, for example, Tscherning (1994).

Spectral combination technique

The least squares spectral combination method (Wenzel, 1982) in our case, yields an optimal combination in the spectral domain of two or more gravity field quantities to obtain second vertical derivative at satellite altitude. The gravity field quantities at a point with spherical polar coordinates (r,ϕ,λ), which are used in the combination are the T_{zz} vertical gravity gradients, computed from torsion balance observations $T_{zz}^{E}(r,\phi,\lambda)$, gravity anomalies $T_{zz}^{G}(r,\phi,\lambda)$, and from a geopotential model $T_{zz}^{M}(r,\phi,\lambda)$. These are combined in a least squares sense to yield the estimate

$$\hat{T}_{zz}(r,\phi,\lambda) = T_{zz}^{M}(r,\phi,\lambda) + T_{zz}^{G}(r,\phi,\lambda) + T_{zz}^{E}(r,\phi,\lambda). \quad (1)$$

Here the torsion balance part with spectral weights w_l can be written by introducing the mean Earth radius R in the form of

$$T_{zz}^{E}(r,\phi,\lambda) = \sum_{l=2}^{\infty}\left(\frac{R}{r}\right)^{l+3} w_l\, \delta T_l^{zz}(\phi,\lambda), \quad (2)$$

where

$$\delta T_l^{zz}(\phi,\lambda) = \frac{1}{4\pi}\frac{2l+1}{l}\iint_\sigma P_l^1(\cos\psi)\left(T_{xz}\cos\alpha^* + T_{yz}\sin\alpha^*\right)d\sigma \tag{3}$$

is the degree l surface spherical harmonic of T_{zz} coming from torsion balance measurements T_{xz}, T_{yz}, which are reduced by the effect of the geopotential model and topography. In the above equation $P_l^1(\cos\psi)$ denotes associated Legendre functions of order 1, ψ is the spherical distance between evaluation and data point, and α^* is the azimuth of the evaluation point measured at the data point. For further explanation of eq. (3) the reader is referred to van Gelderen and Rummel (2001) and Tóth et al (2002). We have followed a two-step process of first converting T_{xz}, T_{yz} data to T_{zz} at the ground level (zero height) using the appropriate kernel function $E_{zz}(r,\psi)$ in

$$T_{zz} = \frac{1}{4\pi}\iint_\sigma E_{zz}(r,\psi)(T_{xz}\cos\alpha^* + T_{yz}\sin\alpha^*)d\sigma\,, \tag{4}$$

and then the formulas of spectral combination with spectral weights were used.

The kernel function $E_{zz}(r,\psi)$ is plotted on Fig 2. We can see that it facilitates the computation of vertical gravity gradients on the ground level well since it is only necessary to integrate the torsion balance data up to a reasonably small integration radius of about 0.3° - 0.5°.

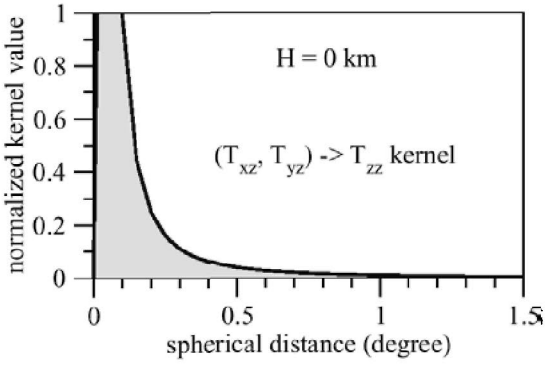


Fig 2. Normalized kernel function $E_{zz}(r,\psi)$ to transfer (T_{xz}, T_{yz}) torsion balance data into T_{zz} vertical gravity gradients at zero height h=0 (cf. eq. 4).

4 Numerical experiments

In the collocation experiment we used all the four gravity gradients (combinations) measured by the torsion balance. To reduce the number of observations we have selected 936 stations of W_{xz}, W_{yz} and 938 stations of $W_{yy\text{-}xx}$, W_{xy} closest to the nodes of a regular geographical grid. Hence altogether 3748 gradients, reduced by the EGM96 gravity field up to maximum degree 360, were used as input. The prediction was taken at 595 points of the simulated GOCE orbit.

We have simulated GOCE gradient observations for 3 months of 2006, between 01.03.2006 and 01.06.2006. Our noise-free GOCE orbit simulation was counting for the gravitational effect of the Earth (using EGM96 model), and the tidal effects of the Sun and the Moon, plus oceanic tides (Schwiderski model was applied). The initial conditions of the satellite orbit were providing similarities to a reliable GOCE scenario, with a mean altitude of 255 km, a mean inclination of 96.5°, and a mean eccentricity of 0.002. Then, the gravity gradients were computed/simulated along this orbit. No noise simulation has been perfomed.

The estimated covariance function of W_{xz}, W_{yz} that was used for the predictions is illustrated in Fig 3.

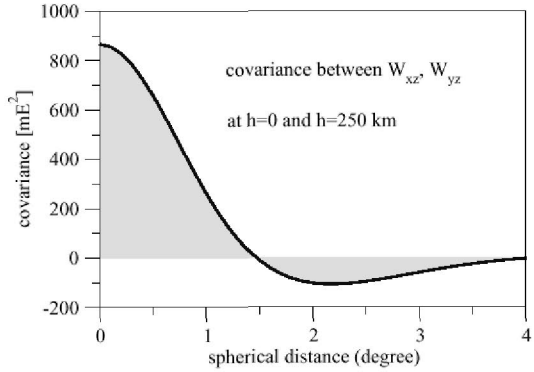


Fig 3. Covariance function of W_{xz}, W_{yz} gravity gradients between zero height of data and the mean prediction level of 250 km.

The simulated gravity gradients were transformed from the local orbital frame to the local North, East, Up frame and compared with the gravity gradients of the EGM96 gravity field model up to degree and order 120. The residuals were very small, the standard deviation was only 0.029 mE. Since we have used the same geopotential model for the collocation prediction, the residuals with respect to the simulated GOCE gravity gradients would be the same. Hence no further comparison was meaningful in this respect.

Table 2. Statistics of predictions at 595 simulated measurement sites. All units are mE.

data set	Min	Max	mean	Std
W_{xz}	-6.9	7.4	0.3	2.7
W_{yz}	-7.0	5.0	-0.1	2.6
$W_{yy\text{-}xx}$	-8.1	4.9	0.0	2.5
W_{xy}	-7.6	4.5	-0.3	2.4

The statistics of the predicted residual quantities can be found in Table 2. The prediction error of these quantities varies between 4-6 mE and on average is about 5.5 mE. We can see that the predicted gravity gradients are very small. This may be attributed partly to the smoothing property of the

collocation but most significantly to the relatively small data collection area of the gradients.

The correlation length of the covariance function of residual gravity gradients at altitude (Fig 3.) is considerably larger than the corresponding correlation length (about 2.5 km) at ground level. This means we would need an extended dataset to reduce the truncation error. The truncation error will be investigated numerically below for the spectral combination method.

The next numerical experiment was made by the spectral combination technique. First, the torsion balance T_{xz}, T_{yz} data were gridded on a 1'×1.5' grid. The extent of this grid was 3°×5.7° in latitude and longitude, respectively. Second, as we mentioned, T_{zz} vertical gravity gradients were computed by integration over the grid at the ground level by eq. (4). For the statistics of these input grids the reader is referred to Table 3. Third, spectral weights were computed using an exponential error covariance function for both gravity anomalies and vertical gravity gradients. For gravity anomalies, which were available over a 3.5°×7° grid, the $C(\psi) = Ae^{-B\psi}$ model was used with parameters A=2 [mGal^2], B=4 [1/°]. This value of B means a 19 km error correlation length of gravity anomalies, the same value that was used for computing the EGG97 European Gravimetric Geoid (Denker and Torge, 1998). The T_{zz} vertical gravity gradients were assigned the same error model but with A=2 [E^2], B=7.7 [1/°]. This value of B means 10 km correlation length, a slightly smaller value than for the gravity anomalies. Since we have no real information on the error covariance function of torsion balance data, the correlation length chosen is only a guessing.

Table 3. Statistics of input grids used for the spectral combination solution. The reference field was the EGM96 geopotential model up to degree and order 360.

data set	Min	Max	Mean	Std
T_{xz} [E]	-132.3	220.3	-0.2	6.9
T_{yz} [E]	-79.5	90.3	-0.0	6.0
T_{zz} [E]	-45.7	67.1	0.1	5.4
Δg [mGal]	-67.5	81.7	-2.8	12.5

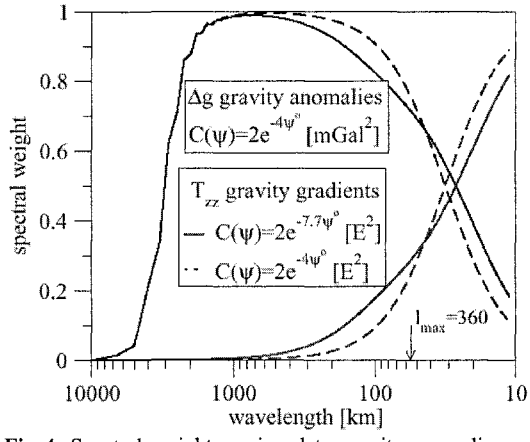


Fig 4. Spectral weights assigned to gravity anomalies and vertical gravity gradients computed from exponential error covariance models (Case B). The dashed lines show spectral weights for the case of equal correlation length of about 0.173° (19 km) of Δg and T_{zz}.

We have computed two different solutions by spectral combination. The first one (Case A) was made using the error degree variances of the EGM96 geopotential model and the ground T_{zz} data only (up to spherical harmonic degree and order 360). The second solution (Case B) was computed by combining gravity information from the EGM96 geopotential model, vertical gravity gradients and gravity anomalies (triple combination). The spectral weights for Case B combination as function of spherical harmonic wavelength can be seen on Fig 4. The dashed lines show the (slightly different) weights for *equal* correlation length of gravity anomalies and gravity gradients. It is remarkable that gravity anomalies dominate the solution between the 30-3000 km wavelengths. This means that a very little contribution is to be expected from surface gravity gradients at the altitude of GOCE. This is confirmed by the results in Table 4. where a 2.3 mE Rms signal was computed in the Case B solution from T_{zz} data.

The spectrally modified kernel functions with the spectral weights assigned for the EGM96 error degree variances and the above two error models can be seen in Fig 5.

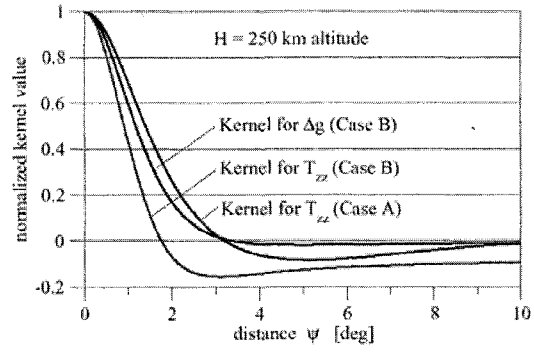


Fig. 5. Integral kernels for the spectral combination solution. The kernels are different for T_{zz} when these data are used alone (Case A) or in combination with Δg (Case B).

To carry out an omission error analysis the truncation error was estimated using the spectral transfer function technique (Wenzel, 1982), adapted to our case. Due to space restrictions we do not give the formulas here, instead we present some plots. The truncation error as function of the truncation cap size for the Case A (T_{zz}-only) and Case B (T_{zz}+Δg) spectral combination solutions is plotted in Figs. 6. and 7. Looking at these plots it can be seen that we have a large truncation error when the torsion balance data are integrated up to about 1°, as is the case with our data. The truncation error is obviously smaller (about 1.5 mE Rms) for Case B solution, but in this case also the signal coming from T_{zz} data is small. This serious truncation error may explain the large differences, discussed below, between this computation and the one computed at

the Insitute für Erdmessung (IfE solution, see Denker, 2002).

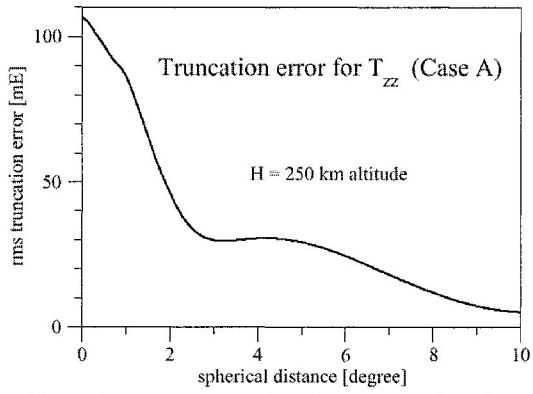


Fig. 6. Truncation error for different cap sizes for the spectral weights assigned to the T_{zz}-only solution (Case A).

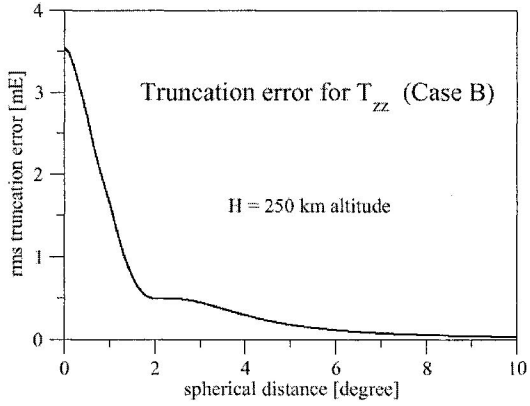


Fig. 7. Truncation error for T_{zz} at different cap sizes for the spectral weights assigned to the $T_{zz}+\Delta g$ solution (Case B).

Finally, a comparison of T_{zz} gravity gradients over Hungary took place at 5'×5' grid points with that of computed at the IfE from EGG97 gravity data by spectral combination technique (Denker, 2002). The differences of the vertical gravity gradients for the $T_{zz}+\Delta g$ solution can be seen on Fig 8. The same differences are shown on Fig 9. for the the T_{zz}-only spectral combination solution.

Table 4. Statistics of $T_{zz} - T_{zz}$(EGM96) residuals computed at H=250 km by spectral combination (Cases A and B) and differences with respect to the solution based on the European gravity data (IfE). All units are mE.

data set	min	max	mean	std
Case A	-4.0	26.0	10.7	8.3
Case A – IfE	-50.8	60.5	32.6	23.6
Case B	-56.2	-0.1	-37.5	39.7
Case B, contribution from T_{zz}	-7.3	2.0	-3.2	2.3
Case B – IfE	-61.9	17.6	-15.5	18.3

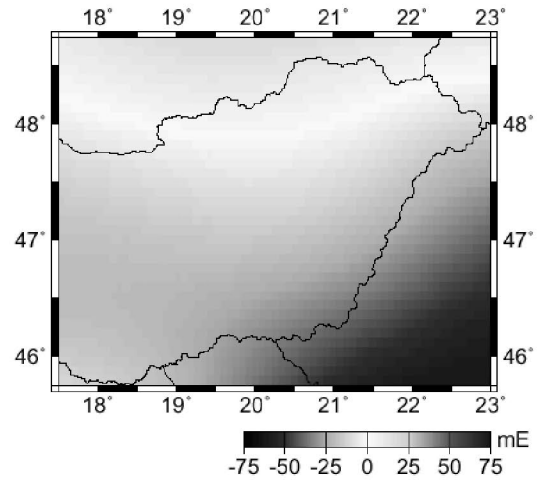


Fig. 8. Differences for T_{zz} computed using Case B ($T_{zz}+\Delta g$) spectral combination solution and that of based on the European gravity data.

The statistics of the results and differences of Case A and Case B solutions with respect to the IfE solution are shown in Table 4. It can be seen that the residual values computed in this experiment are smaller (especially for Case A) and go zero outside Hungary. This also indicates significant truncation errors in the results. The two spectral combination solutions (Case A and B) show big differences. The solution with gravity anomalies (Case B) agrees better with the IfE solution – at least they show a similar low over Hungary in the residual vertical gravity gradients. But even in this case we expect large truncation errors in the integration of gravity anomalies since our gravity data are extended over a 3.5°×7° area only.

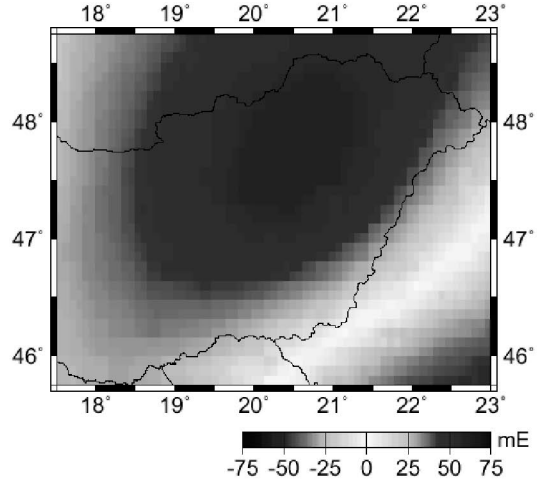


Fig. 9. Differences for T_{zz} computed using the Case A (T_{zz}-only) spectral combination solution and that of based on the European gravity data

5 Conclusions, recommendations

The most important conclusion that can be drawn from the above study is that the torsion balance data in their present data coverage fail to perform a meaningful calibration/validation for the (relatively) high altitude of the GOCE satellite. This is mainly because we found the ommission error at this altitude very serious for torsion balance gravity gradients. The comparisons with the IfE solution supports our conclusion in this respect. Moreover, if one considers the anticipated accuracy of the GOCE gravity gradients of about 4 mE/√Hz for the 5-100 mHz measurement bandwidth, 1 mE accuracy of the calibration/validation gradients is required. As we have seen, the truncation errors exceeds this accuracy by at least a factor of 10.

Another factor in explaining the differences is the supplementing character of ground and space gravity gradients. Space gravity gradient data correlate best with ground gravity anomalies and/or geoid heights, not with ground gravity gradients. This is due to the smoothing effect of upward continuation. Therefore we conclude one may use torsion balance data only as a supplementary information giving a few mE signal , in addition to a gravity anomaly dataset of good data extent as it was the case of computations by Denker (2002). This combination would be a meaningful one.

On the other hand, torsion balance data are still valuable in checking ground gravity field information. As it can be seen from the spectral weights in Fig 4, below of about a 30 km wavelength torsion balance data may provide better gravity field modeling than gravity anomalies. Therefore it would be interesting to use these data in local geoid determination using the spectral combination technique as discussed in this paper above. In the context of GOCE, torsion balance data can support gravity field determination in the short wavelengths, due to their above mentioned supplementary character.

Acknowledgements

The GOCE simulation has been performed by the orbit integrator at the IAPG, Technical University of Munich, developed by M. Wermuth, B. Frommknecht and P. Steigenberger. The support of the Greek-Hungarian Interngovernmental Scientific and Technology Fund is also acknowledged.

References

Arabelos, D., and C.C. Tscherning (1998). Calibration of satellite gradiometer data aided by ground gravity data. Journal of Geodesy, 72:617-625.

Denker, H. and Torge, W. (1998). The European gravimetric quasigeoid EGG97 – An IAG supported continental enterprise. In: R. Forsberg et al. (eds.), Geodesy on the Move – Gravity, Geoid, Geodynamics, and Antarctica, IAG Symp. Proceed., Springer Verlag, Berlin, Heidelberg, 119:249-254.

Denker, H. (2002). Computation of Gravity Gradients for Europe for Calibration/Validation of GOCE Data. Proceedings of the Gravity and Geoid 2002 3rd Meeting of the Internat. Gravity and Geoid Comm., Aug. 26-30, Thessaloniki, 2002.

Hein, G.W. (1981). Untersuchungen zur terrestrischen Schweregradiometrie. Dissertationen Reihe C 264, Deutsche Geodätische Kommission, München.

Tóth Gy., Rózsa Sz., Ádám J. and Tziavos, I.N. (2002). Gravity field modeling by torsion balance data – a case study in Hungary. In Ádám J. and Schwarz, K-P., editors, Vistas for Geodesy in the New Millennium, Vol 125. IAG Symposia, pp 193-198.

Tscherning, C.C. (1994). Geoid determination by Least-squares Collocation using GRAVSOFT. In: International School for the Determination and Use of the Geoid, Milan, October 10-15, 1994, pp 135-164.

van Gelderen, M. and Rummel, R. (2001). The solution of the general boundary value problem by least squares. Journal of Geodesy, 75:1-11.

Wenzel, H.G. (1982). Geoid computation by least squares spectral combination using integral formulas. Proceed. Of the General Meeting of the IAG, Tokyo, May 7-15, pp 438-453.

SYMPOSIUM G03

Determination of the Gravity Field

M.G. Sideris

Repeated Measurements of Gravity with the Absolute Gravimeter FG5 #210 at Matsushiro, Japan and Comparison with the Superconducting Gravimeter T011

Y. Imanishi
Ocean Research Institute, The University of Tokyo, 1-15-1, Minamidai, Nakano, Tokyo 164-8639, Japan

T. Higashi, Y. Fukuda
Department of Geophysics, Graduate School of Science, Kyoto University, Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8224, Japan

K. Kokubo
Matsushiro Seismological Observatory, Japan Meteorological Agency, 3711 Nishijo, Matsushiro, Nagano, Nagano 381-1232, Japan

Abstract. Absolute gravity measurements have been repeatedly made with the FG5 #210 at the superconducting gravimeter station in Matsushiro, Japan. Four out of the five experiments so far performed were successful. The data from the absolute gravimeter are used to calibrate the instrumental sensitivity and drift of the superconducting gravimeter T011 with good results. Drift of the absolute gravimeter is investigated at the same time. Long-term gravity changes are discussed by collocating the data from the absolute and superconducting gravimeters.

Keywords: absolute gravimeter, superconducting gravimeter, calibration, drift, long-term gravity changes

1 Introduction

To make absolute gravity measurements at a superconducting gravimeter (SG) station is useful for various purposes. Comparison of the data from the absolute gravimeter and the SG is the well-established method of calibrating the instrumental sensitivity of the latter (e.g. Hinderer et al., 1991). At the same time, it also provides an opportunity of checking the absolute gravity measurements against the SG data (Imanishi et al., 2002). If the absolute gravity measurements are performed repeatedly at the same SG station, the small but non-negligible instrumental drift of the SG as well as steps in gravity can be calibrated against the absolute measurement results. Since the instrumental drift of the SG is known to be very close to either linear or exponential with respect to time, this in turn may provide information on the status of each experiment of absolute gravity measurements. Normally the SG is installed at such a place where artificial noise level is low and environmental condition is stable, absolute gravity data obtained at the SG station is generally expected to be of high quality. This will result in a high-quality dataset of long-term gravity changes observed at the station.

Comparison of the FG5 absolute gravimeter manufactured by Micro-g Solutions and the Superconducting Gravity Meter manufactured by GWR Instruments has been already performed on a routine basis by the groups owning both instruments (Amalvict et al., 2001). Such experiments are extremely important because they will provide ground truths for the satellite gravity missions such as GRACE and CHAMP.

In this paper, we report our recent results of absolute gravity measurements at the SG station in Matsushiro, Japan (Imanishi, 2000). The FG5 absolute gravimeter (serial number #210) is used throughout the experiments. Since this station is in a very stable environment and the ground noise level is relatively low, it is expected that a consistent data set of temporal gravity changes would be obtained by collocating the absolute gravity measurements and the continuous gravity

observation with the SG. Although this station is not in a tectonically active region, long-term changes in gravity could be detected if crustal deformation is under way at a steady rate. This is the first systematic challenge of high-precision gravity monitoring with the combination of the absolute gravimeter and the superconducting gravimeter in Japan.

2 Absolute Gravity Measurements

The gravity station where our experiments of absolute gravity measurements were performed belongs to the Matsushiro Seismological Observatory, Japan Meteorological Agency. The measurement site is a permanent gravity station called the Fundamental Gravity Station (FGS) built by Geographical Survey Institute (Yamamoto et al., 1998). This is located in the gravimeter room inside the major tunnel of the observatory, about 2 m apart from the SG T011.

The first measurement of absolute gravity at the Matsushiro SG station took place in 1996 (Okubo et al., 1997). This was followed by another measurement in 1997 made by Geographical Survey Institute. Analysis and interpretation of the results of these early measurements will be presented elsewhere. Here we will focus on the experiments which were performed after 2000.

Table 1 lists the parameters and the results for the five experiments of absolute gravity measurements performed from 2000 through 2002. This includes the experiment in August 2001 which was not successful because of a failure of the FG5 gravimeter. The experiments have been performed on a semi-routine basis at approximately half a year interval. The same gravimeter (FG5 #210) and measurement procedure were used throughout the experiments. The measurement period spanned typically from one to four weeks. The interval of each drop (sampling interval) was 15 s. Normally, the measurements consisted of hourly 'sets' each of which consisted of 160 drops within 40 minutes. The standard deviation of each drop was typically less than (5-10) $\times$ 10^{-8} ms^{-2}, and was highly dependent on the ground noise level.

In each experiment except August 2001, the statistical distribution of measured gravity values was very close to a normal distribution (Fig. 1), as typically observed when the measurements were performed normally. The mean and the standard deviation of the distribution varied from experiment to experiment, as shown in Table 1. The mean gravity values have been corrected for the measurement height so that they refer to the top surface of the FGS metal plate. Thanks to the quiet and stable condition of the station, relatively small values of the standard deviation for each drop have been obtained. In particular, the April 2000 result is characterized by a small standard deviation and a large number of available data, and may be one of the best results of absolute gravity measurements ever conducted in Japan. Here we have revised the gravity value for the April 2000 experiment reported in Imanishi et al. (2002) regarding the correction of the polar motion effect.

The mean gravity values obtained from the four successful experiments need not to agree exactly with each other, because there may have been real changes in gravity acceleration. However, the result for the December 2001 experiment looks significantly different from the others, even if we take into account possible gravity changes of natural origin. There are evidences showing that, in the December 2001 experiment, the frequency of the rubidium clock used with the FG5 may have been lower than the prescribed value (10 MHz) by some parts in 10^9. In general, a relative error in the frequency, $\Delta f/f$, results in an erroneous standard of time and therefore an error in the measured gravity

Table 1. Summary of the experiments of absolute gravity measurements performed at Matsushiro SG station with FG5 #210. Mean gravity and standard deviation (SD) refer to the mean and the standard deviation of the statistical distribution of the raw measurements. Scale factors of the SG are estimated with two different methods of analysis, the simple least squares method (LSQ) and the modified least squares method (MLSQ).

Date	# day	# data	Mean Gravity (ms^{-2})	SD (10^{-9} ms^{-2})	Scale Factor (10^{-9} $ms^{-2}V^{-1}$) LSQ	MLSQ
April 2000	27	98200	9.797729815	72	928.01 ± 0.34	928.79 ± 0.36
December 2000	17	46800	9.797729769	90	928.49 ± 0.57	928.61 ± 0.58
August 2001	-	-	-	-	-	-
December 2001	9	30100	9.797729946	102	928.58 ± 0.69	927.76 ± 0.80
August 2002	3	12600	9.797729836	68	931.12 ± 1.14	931.45 ± 1.13

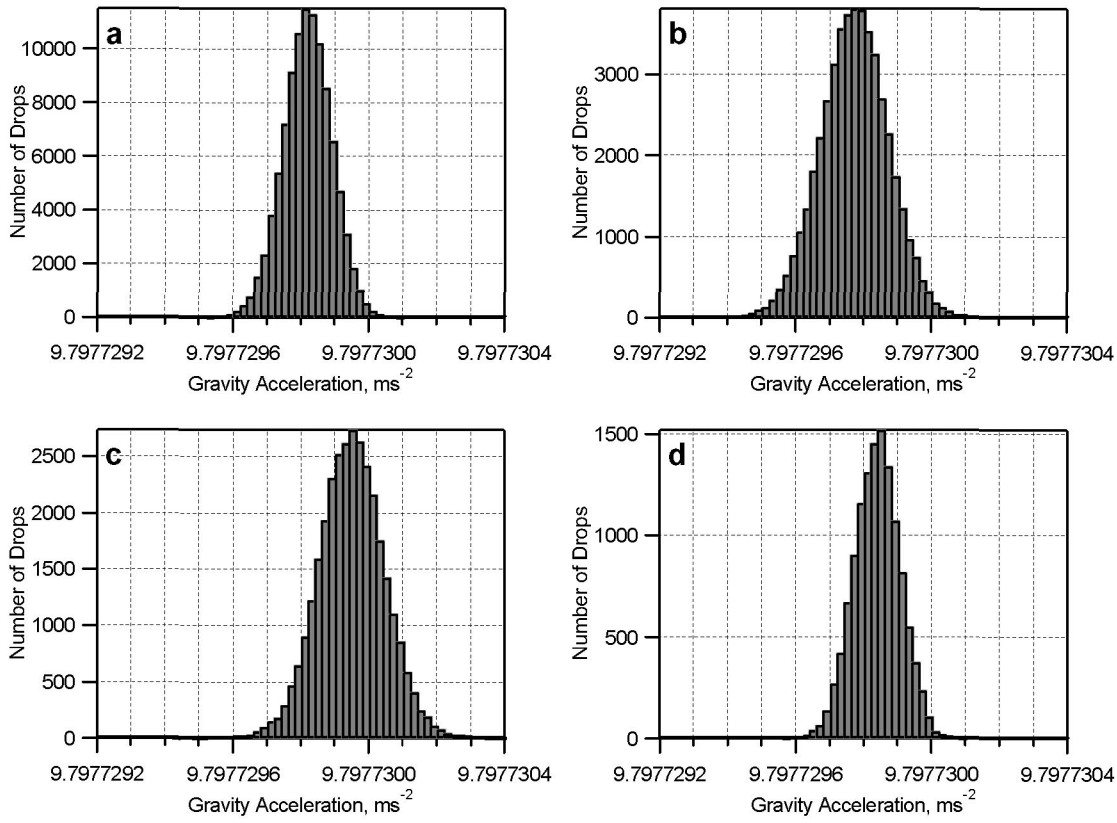


Fig. 1. Histograms of the absolute gravity values measured in each experiment. **a** April 2000 **b** December 2000 **c** December 2001 **d** August 2002.

acceleration as large as $\Delta g/g \sim 2\Delta f/f$. Therefore, in the December 2001 experiment, the estimate of the gravity acceleration may have been larger than the true value by (5-10) $\times$ 10^{-8} ms^{-2}. Unfortunately, the rubidium clock used in this experiment was not calibrated precisely against other frequency standards, and so we can not make a precise correction to the measured value. In this paper, we regard the absolute gravity estimate from the December 2001 experiment as an outlier and do not use it in the discussion of the mean value of gravity. After this experience, calibration of the rubidium clock against another well-calibrated rubidium clock or a cesium clock has been included in our observation routine.

3 Calibration of the Superconducting Gravimeter and Stability of Absolute Gravity Measurements

For each experiment, absolute gravity data are used to calibrate the instrumental sensitivity of the SG T011. The clock problem of the December 2001 experiment mentioned above does no harm on the calibration, because only the relative changes of gravity matter in the comparison with the SG data.

Normally, one fits a straight line to the absolute gravity data (y) against the SG data (x) as in $y = a\,x + b$ to estimate the scale factor a of the SG. Table 1 lists the (absolute) scale factors obtained in this way for the four experiments. Except the August 2002 experiment which spanned for only 3 days, the three experiments have given practically the same results within the estimation errors.

Imanishi et al. (2002) reported a problem in the calibration of the SG with the absolute gravity measurements which was observed in the April 2000 experiment. In the least squares analysis of the data, the constant term b was found to be variable with time, and therefore the estimate of the scale

factor a was considered to be biased. An original method of data analysis was then developed based on the Bayesian statistics by incorporating a trend part into the least squares method, so that slow irregular changes in the constant term could be taken into account. In that method, the basic equation of the model was extended to the form $y_n = a\,x_n + b_n$, where n is the index of the data samples. Here b is no longer a constant but variable with time, representing an irregular trend in the offset term. The model is fit to the data in the least squares sense but with an additional condition that the changes in b should be smooth with time. Thus, the weighted residual (with weight w_n):

$$S = \sum_n w_n (y_n - a x_n - b_n)^2 + \alpha^2 \sum_n (b_n - 2b_{n-1} + b_{n-2})^2$$

is minimized with respect to the unknown parameters a and b_n. Here α^2 is the hyperparameter which specifies the smoothness of the trend part. α^2 is chosen so that an information-theoretical criterion called Akaike Bayesian Information Criterion (ABIC) is minimized. From the analysis of the data with this method, a possible "drift" in the absolute gravity measurements was detected and a supposedly unbiased estimate of the scale factor was obtained. Details on this modification of the least squares method is given in Imanishi et al. (2002).

Here we have analyzed the data from the other three experiments with the modified least squares method to examine if a similar phenomenon is observed. Fig. 2 shows the estimated trend parts which correspond to temporal changes of the constant term. The result for the April 2000 experiment is also reproduced here. The instrumental drift of the SG has been subtracted beforehand using the parameters obtained from many years of operation. These results have been obtained with different hyperparameters of the model. Therefore, we can not compare directly the relatively short period components in these results. In addition, the available length of data differs significantly between these experiments. Nevertheless, we may say that the trend parts for the latter three experiments (Fig. 2b, c, d) do not look quite different from that for the April 2000 experiment (Fig. 2a). The trend for these experiments in shorter duration would have indicated similar or larger changes in a 27-day long experiment. We may tentatively conclude that, as far as our results are concerned, the drift of the absolute gravity measurements is not special to the April 2000 experiment but observed universally.

The above conclusion leads to the necessity of adopting the scale factors of the SG obtained from the modified least squares method, instead of those from the simple least squares method. Table 1 compares the results for the four experiments obtained with the two methods. Depending on the drift in each experiment, the two methods may or may not give significantly different results. The modified least squares method tends to give a slightly larger estimation error than the simple least squares method. When we compare the estimates for the four experiments obtained by the modified least squares method, the results for the December

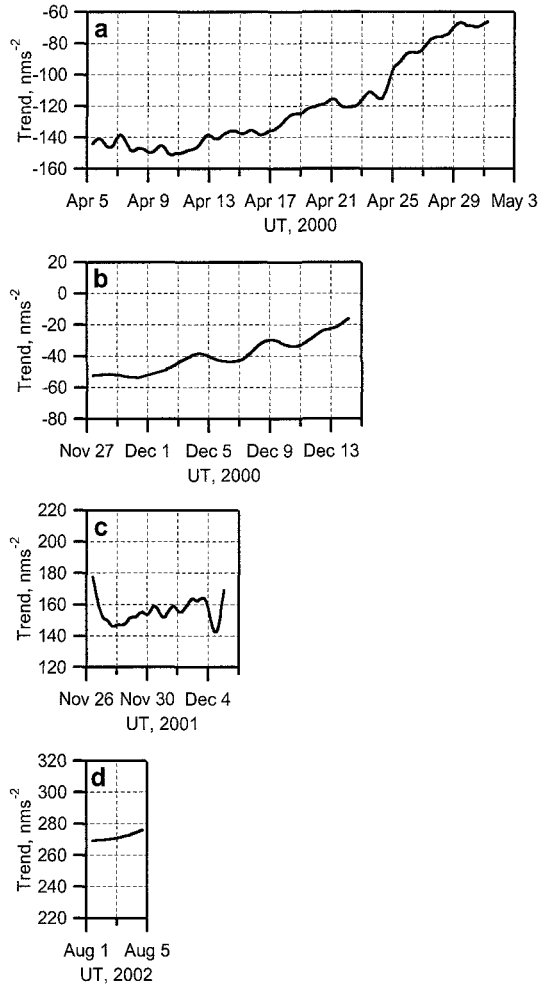


Fig. 2a, b, c, d. Irregular changes of the constant term in the least squares analysis of the absolute gravity measurements against the superconducting gravimeter data. **a** April 2000 **b** December 2000 **c** December 2001 **d** August 2002 experiment.

2001 and August 2002 experiments look significantly different from the others. In the December 2001 experiment, the trend part (Fig. 2c) has a large correlation with the tidal signal, which should result in a large bias in the scale factor. This must be the reason for the large difference between the scale factors estimated by the two analysis methods. The August 2002 experiment may be too short to give a reliable calibration result. Considering these, we conclude that the true scale factor lies between or around the results for the two experiments performed in 2000.

4 Long-term Gravity Changes

In the period from the beginning of 2000 to the end of 2002, the SG T011 at Matsushiro was in a stable condition and was producing continuous recordings of gravity changes almost free from gaps and large steps. Although there are small steps of instrumental origin in the data of the year 2000, they appear always in a pair of positive and negative steps of the same magnitude (~ 1 mV ~ 10^{-9} ms^{-2}) and do not have a serious effect on the estimation of the gravity trend in a long time span. This enables us to compare the discrete results of absolute gravity measurements with the continuous trend from the SG to investigate how the long-term gravity changes at the station are recorded by the two instruments.

Fig. 3a shows the residual gravity series from the SG as well as the results of the four experiments of absolute gravity measurements. The tides (solid earth + ocean), the effects of atmospheric pressure and polar motion have been removed from the both data. For the atmospheric effect, a local correction using a single coefficient of admittance (3.2×10^{-9} $ms^{-2}hPa^{-1}$) has been applied. Here the instrumental drift of the SG has been retained in the data. The difference between the gravity values from the two instruments, shown in Fig. 3b, can be used to estimate the instrumental drift of the SG. The instrumental drift of the SG T011 has been known to be very close to a decaying exponential function of time with time constant of about 7 years (Imanishi, 2000). After more than 11 years of operation since the first installation at the older station Kakioka (Imanishi and Segawa, 1998), the exponential drift has already become almost indistinguishable from a linear function of time. So we assume a linear drift to be fit to the data. Excluding the data for the December 2001 experiment, a weighted least squares fit yields a drift rate of $(1.8 \pm 0.4) \times 10^{-7}$ $ms^{-2}a^{-1}$ (gravity increasing). Thus, the empirical drift rate of 2×10^{-7} $ms^{-2}a^{-1}$ so far estimated solely from the SG data has been verified.

Removing the instrumental drift from the series in Fig. 3a, we obtain the final estimate of the residual gravity from the SG, as shown in Fig. 3c. We believe that this is very close to the real gravity changes at the Matsushiro station. Although there are still discrepancies between the two gravimeters, the discrepancies are $(2 - 4) \times 10^{-8}$ ms^{-2}, comparable to the precision of the FG5 absolute gravimeter.

The peak-to-peak magnitude of the residual

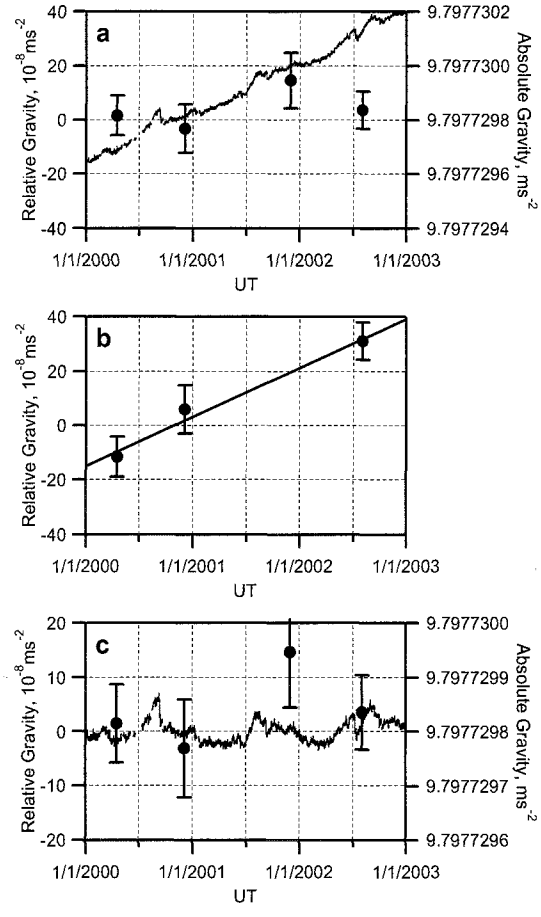


Fig. 3a, b, c. Comparison of long-term gravity changes recorded by the superconducting gravimeter and the absolute gravimeter. **a** Residual gravity from the superconducting gravimeter (solid line, left axis) compared with the absolute gravity measurements (dots, right axis). **b** Difference between the gravity values from the two gravimeters, approximated by a linear function of time. **c** Same as **a** but with the instrumental drift removed. Note that the vertical axis is expanded.

gravity changes is 1.0×10^{-7} ms^{-2}. Seasonal changes are evident in the data; the gravity increases from July to September in every year. This is very likely to be associated with hydrological effects, mainly originating from local rainfall. At Matsushiro, we have more rainfall in summer than in other seasons. Although we observe gravity decrease after rainfall (Imanishi, 2000), instead of increase, in the short time scale, the effect of rainfall can have a reversed sign in a longer time scale. Unfortunately, the relative gravity values were close to the mean when the four (successful) experiments of absolute gravity measurements were made. A few more times of absolute gravity measurements in those seasons when large signals of the local hydrological effects are expected would verify the relation between the rainfall and the observed gravity changes. For the present, we speculate that the residual gravity changes from the SG would be reduced significantly if the hydrological effects are appropriately taken into account.

Finally, no clear evidences for secular changes in gravity have been detected from the data so far obtained. Although we do not expect a large rate of gravity changes due to crustal deformation at Matsushiro, it will be interesting to keep track of the secular changes in gravity with such a high precision that would be difficult to achieve at other SG stations in Japan.

5 Conclusions

Repeated measurements of absolute gravity at Matsushiro have produced promising results for various applications. In particular, comparison of the absolute and the superconducting gravimeters at Matsushiro has proved to be very efficient in monitoring the long-term gravity changes. We plan to continue our project in the future so that more definitive conclusions can be drawn on the hydrological effects on gravity. Also, installation of a permanent station for interferometric GPS is planned so that the gravity changes due to crustal deformation can be resolved from the other effects on gravity.

References

Amalvict M, Hinderer J, Boy JP, Gegout P (2001) A three year comparison between a superconducting gravimeter (GWR C026) and an absolute gravimeter (FG5#206) in Strasbourg (France). J Geod Soc Japan 47: 334-340

Hinderer J, Florsch N, Mäkinen J, Legros H, Faller JE (1991) On the calibration of a superconducting gravimeter using absolute gravity measurements. Geophys J Int 106: 491-497

Imanishi Y (2000) Present status of SG T011 at Matsushiro, Japan. In: Ducarme, B and Barthélemy, J (eds) Proceedings of the Workshop: High Precision Gravity Measurements with Application to Geodynamics and Second GGP Workhop, Cahiers du Centre Européen de Géodynamique et de Séismologie, 17, 97-102

Imanishi Y, Higashi T, Fukuda Y (2002) Calibration of the superconducting gravimeter T011 by parallel observation with the absolute gravimeter FG5 #210 - a Bayesian approach. Geophys J Int 151: 867-878

Imanishi Y, Segawa J (1998) Earth tides and residuals observed by a superconducting gravimeter at Kakioka, Japan. In: Ducarme B, Pâquet P (eds) Proceedings of the 13th International Symposium on Earth Tides, 575-582

Okubo S, Yoshida S, Sato T, Tamura Y, Imanishi Y (1997) Verifying the precision of a new generation absolute gravimeter FG5 - comparison with superconducting gravimeters and detection of oceanic loading tide. Geophys Res Lett 24: 489-492

Yamamoto H, Nishimura O, Fujiwara S (1998) Establishment of Japan Gravity Standardization Net 1996. Bulletin of the Geographical Survey Institute 44: 1-10

APPLICATION OF GRAVIMETRY BY HELICOPTER TO IDENTIFY MARINE ACTIVE FAULTS AND IMPROVE ACCURACY OF GEOID AT COASTAL ZONES

Jiro Segawa [1], John Joseph E. [2], Eiji Nakayama [3], Vijay Kumar K. [1], Shigekazu Kusumoto [1], Takeo Ito [3] Seiichi Sekizaki [3], Takemi Ishihara [4], and Masao Komazawa [4]

[1] School of Marine Science and Technology, Tokai University, 3-20-1 Shimizu Orido, Shizuoka 424-8610 Japan.
E-mail: jsegawa@scc.u-tokai.ac.jp
[2] Geology and Geophysics, School of Earth and Environmental Sciences, the University of Adelaide, SA 5005, Australia
[3] Aero Asahi Corp., 1-18-1 Shin-Sayama, Sayama, Saitama 350-1331 Japan
[4] Geological Survey of Japan, 1-1-1 C#7 Higashi, Tsukuba, Ibaraki 305-8567, Japan

ABSTRACT

We started the development of gravimetry using helicopter in 1998 and attained a preliminary success in gravity measurement in 2000. For this purpose we have manufactured a new gravimeter (SEGAWA Model) with a servo accelerometer as a gravity sensor and an optical fibre gyro to keep vertical. Our success owes mainly to precise positioning of helicopter, and delicate correction for the effect of helicopter's horizontal acceleration. We have so far conducted test and/or practical measurements by helicopter (mostly using Bell 412). The advantages of a helicopter over a fixed-wing aircraft in gravity measurement is the high resolution of gravity anomaly obtained because of low altitude flight and low and stable flight speed. The main objective of the measurements is to get continuous profiles of gravity anomalies in the land-to-sea boundary zones, so that gravity void or data discrepancy across coastal zones may be removed and that the active seismic faults running across coasts, which are reflected on gravity anomalies, may be retraced as far as the sea floor. In this paper two results have been selectively described: One is the result of measurement conducted in April 2000 along west-to-east tracks from Saitama to Kashima-Nada Sea, which has disclosed inconsistency between land and marine gravity. Another is the result of survey along the east-to-west tracks in the Enshu-Nada Sea (Tokai area), which shows the changes of free air anomaly and/or gravity disturbance from the coast to the sea floor associated with active seismic faults.

Key Words: helicopter gravimetry, gravity disturbance, land-to-sea consistency of gravity, sea floor seismic fault

1. Introduction

We started the development of a new gravimeter for helicopter use in 1998 (Segawa et al., 2000; Segawa et al., 2001; Segawa et al., 2002). The gravimeter which is preliminarily named SEGAWA model consists of a servo accelerometer type gravity sensor, a stabilized platform controlled by an optical fibre gyro, a set of carrier phase interferometric GPS positioning system (post processing), a DGPS system (real time) based on pseudo-range measurement and its correction, and a data processor. The drift rate of gravity sensor is 0.5 to 1 mgal/day and it is almost linear with time when the earth tide is removed (Segawa et al., 2000). The vertical of gravity sensor is kept by an optical fiber gyro which indicates the vertical on the basis of position (latitude and longitude) provided by DGPS system. Since the optical fiber gyro used has a 3 minutes of arc maximum ambiguity in verticality we have to consider the error caused by horizontal acceleration of helicopter. To solve this problem we have installed a pair of horizontal accelerometers on the stabilized platform and estimated the offset of the vertical by making comparison between the measured and the calculated (from GPS) horizontal acceleration (Refer to Peters and Brozena, 1995). The correction for horizontal acceleration amounts usually to 20 to 150 mgal, depending on atmospheric conditions and/or navigation technique of the pilot. Under a favorable condition we can correct for the horizontal acceleration effect with the accuracy of 1 to 2 mgal.

It is important that the gravity anomaly obtained

Fig.1 Gravity survey points and tracks in and around the central part of Honshu, Japan after Kuroishi (2002). Grey and black areas indicate the places where there are dense gravity data. The blank areas are gravity void. Blank zones at sea along the coast are 10 to 20 nautical miles in width occupying areas which cannot be neglected.

from airborne gravimetry is not free air gravity anomaly but gravity disturbance because gravity is measured on a plane defined by ellipsoidal height, not on the geoid. Therefore, when the airborne gravity measurement is to be compared with ship-borne gravity measurement we have to take the undulation of geoid into account where it is necessary. The free air gravity anomaly on the geoid is obtained by subtracting the normal gravity from the measured gravity after free air reduction based on the elevation above geoid. The gravity disturbance at a certain altitude is the difference between the measured gravity at altitude and the normal gravity after free air reduction based on the ellipsoidal height. Since the relative height between geoid and the flight level changes due to the geoid undulation the gravity disturbamce and the free air gravity anomaly do not always behave in parallel.

The advantage of helicopter-mounted gravimeter over the gravimeter on a fixed-wing aircraft is the capability of measurement at low altitude and low speed, resulting in a better resolution of gravity anomalies. Since our present target is to make surveys at flat fields near the coastal zones we can usually fly at the altitude of 2000 ft with a speed from 45 knots to 90 knots (83 km/h to 166 km/h). The resolution of gravity measurement depends on the flight altitude, speed and noise filter. It is evident that the lower the altitude the higher the resolution is. As for the flight speed it is desirable to be as low as possible while considering the stability of navigation and the efficiency of survey. The digital filter used is either the Gaussian type or the Cosine type lowpass filter. The half width of the digital filter is 100 to 150 sec. This length of filter will admit signals with the minimum period of 2 min to 3 min, or the minimum wavelength

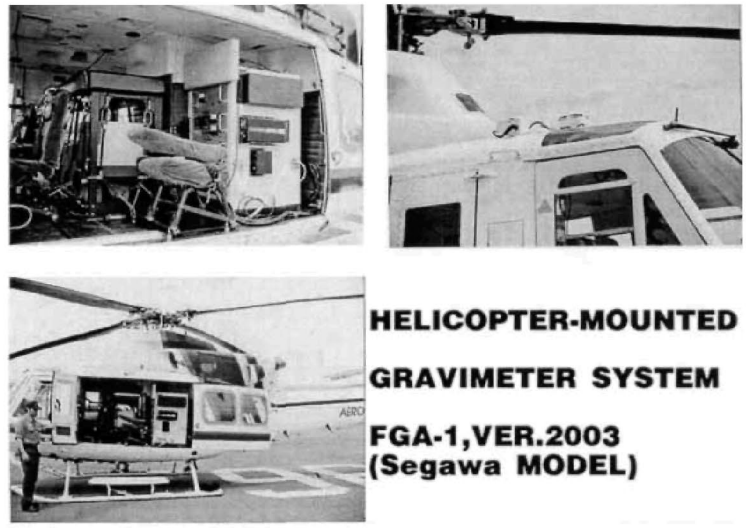


Fig.2 Installation of helicopter-mounted gravimeter system on Bell 412. Upper left shows the gravity sensor/gyro unit and the control unit installed in the passengers' room . Upper right shows two GPS antennas installed on the roof of the helicopter. Lower left shows the outer view of Bell 412.

of gravity anomaly 3 to 4 km for the case of the flight speed 45 knots. In the case of usage of a fixed-wing aircraft it is difficult to fly at a low speed so that the wavelength of detectable signals is forced to be longer. As an example the final accuracy of gravity measurement is shown in Segawa et al. (2002), where the repeatability of gravity measurement is represented for the case of the measurement in April 2000 at the Kashima-Nada Sea. The data from this measurement shows that the bias difference and the standard deviation between the round track measurements is 0.5 mgal and 1.5 mgal, respectively.

2. Missions of gravity measurement

A problem we are faced with at present is the gravity void spreading along the coastal zones of the Japanese Islands. The areas void of gravity data are shown by blank zones in Figure 1 (Kuroishi, 2002). The gravity void found along the coasts not only deteriorates geoid estimation but also disables us from judging the quality of marine gravity measurements or handling gravity data continuously from land to sea. Under this situation we have planned helicopter gravity measurement to get objective data with which to fill the gravity void as well as to evaluate bias errors of marine data. Another application of helicopter gravity measurement is to track active seismic faults gravimetrically identified on land towards the sea. As explained later in detail we have extremely dense gravity measurements on land in the Japanese Islands, which enable us to draw gravity anomaly contours with the interval of 0.5 mgal (Gravity Research Group in Southwest Japan, 1994). This sort of gravity map reveals significant linear patterns of gravity anomalies coinciding with geologically identified trend of active seismic faults.

In this paper we will report the results from two missions; one about the bias error with the marine gravity measurements and the other about possibility of tracing active seismic faults on land over to the sea floor on the continental shelves.

3. Gravimeter assembly

Figure 2 shows the outer view of our helicopter-mounted gravimeter installed on BELL 412 helicopter. This helicopter can accommodate 13 passengers and fly with the maximum speed of 120 knots for 2 hours. The gravimeter is assembled as three units, that is, gravity

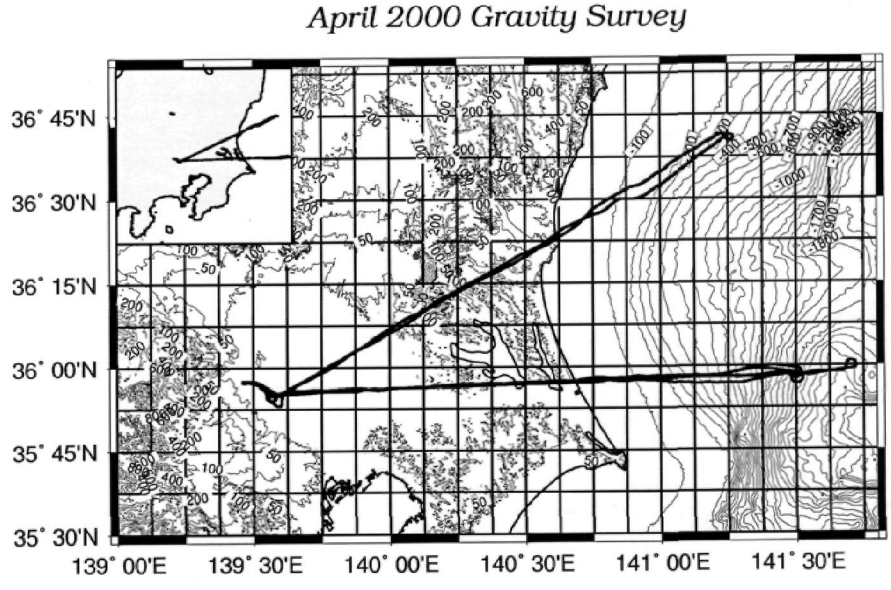


Fig.3 Tracks of helicopter gravity measurements in April 2000 from Saitama to Kashima-Nada Sea via Ibaraki. In the text are described the results from the west-to-east tracks along 35° 55′ N only. Four round flights were tried along these tracks at the altitude of 2000 ft, but two of them are selectively explained in this paper. The tracks span about 200 km.

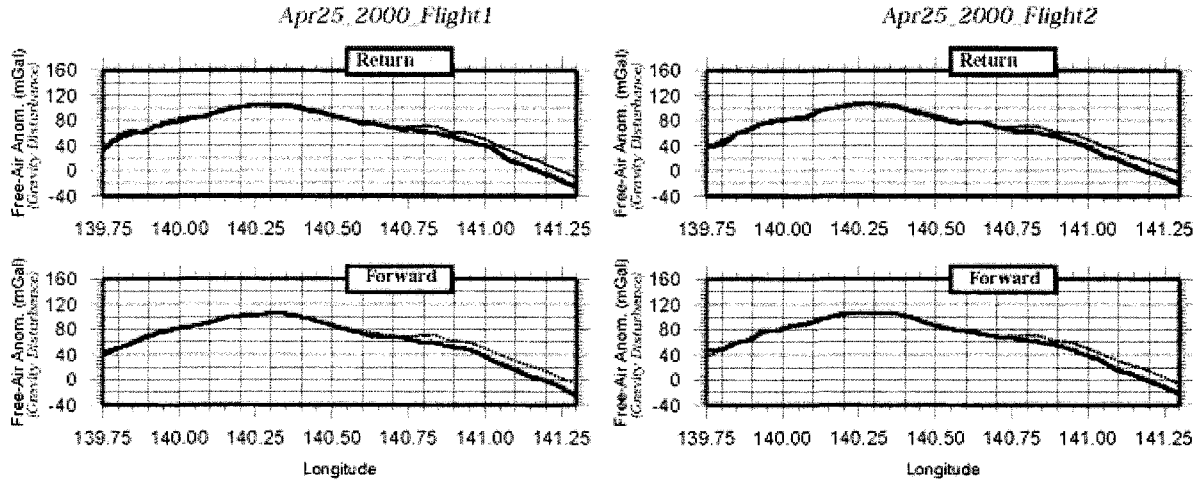


Fig.4 Profiles of gravity anomaly along the tracks in Fig.3. Lower left is the profile on the outgoing track in the morning on 25 April 2000, upper left the profile on the return track of the same flight (flight 1). Lower right and upper right profiles (flight 2) are similar to the left profiles except that its flight was in the afternoon. What we should point out here is that upward continued ground/sea truth data (dotted curves) detach from the profiles of gravity disturbance by helicopter (solid curves) east of the coastal line of Kashima-Nada Sea (140.70 ° E). The difference between gravity disturbances and ground/sea truth free air anomaly data is approximately 15 mgal. We believe this is due to the bias error involved in the old surface ship data.

sensor/gyro unit, control/data recording unit and positioning unit. The former two units are installed in the passengers' room and GPS antennas of the positioning unit are installed outside on the helicopter's roof. Detailed description of this gravimeter can be referred to Segawa et al. (2000).

4. Finding inconsistency between land and marine gravity (mission 1)

Marine gravity around Japan has been measured by surface ship gravimeters since 1960's. Although nobody doubts land gravity data as long as 1 mgal accuracy is concerned the marine gravity data are suspected to involve larger errors mainly because of positioning errors or other unfavorable conditions. But, nothing has so far been made to examine the plausibility of old data and improve them if there should be problems. So, we have made it our task to check air gravity measurement against land gravity data, and, if the air measurement is good within land area, then it is extended and compared with marine gravity data. It is important in this procedure that the helicopter measures land and sea gravity at a time and that comparison be made using a proper upward or downward continuation techniques.

The helicopter gravity measurement in April 2000 was conducted under this principle, and we selected the tracks of measurement as seen in Figure 3. The area of survey is from Saitama, Ibaraki offshore to the Kashima-Nada Sea, Japan. During the measurements we conducted round flights three times on the same track and one round flight on a different track. The west-to-east tracks of Figure 3 are approximately along 35° 55'N. The track passes Kawagoe Heliport of Saitama eastward to Kashima-Nada Sea via Ibaraki. Another flight is in the southwest-to-northeast direction. The length of the tracks of measurement is approximately 200 km each. In this paper the gravity disturbances on the level of 2000 ft obtained by two flights on the west-to-east tracks (Flight 1 and Flight 2) are shown (Figure 4) together with the ground/sea truth free air gravity anomalies. The ground/sea truth data are upward continued to the level of 2000 ft.

The four profiles of free air gravity anomaly at the altitude of 2000 ft are compared with ground/sea truth data which have been upward continued to the level of 2000 ft. The left two profiles in Figure 5 are those obtained by the first flight (onward and backward) in the morning 25 April, while the right two profiles are those obtained by the second flight in the afternoon on the same day. What is worth mentioning is that the ground/sea truth data disagrees with all the helicopter gravity in the area from 140.70°E eastward. This longitude is the very point of the coast which separates Ibaraki from Kashima-Nada Sea. This result shows that the marine gravity at Kashima-Nada Sea is by about 15 mgal larger than the gravity determined by helicopter. We believe that, since the helicopter

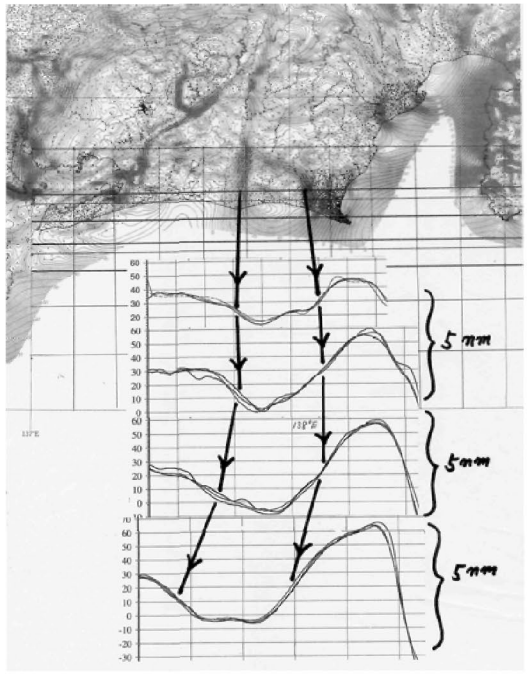


Fig.5 A very dense Bouguer anomaly contour map (0.5 mgal contour interval) after Gravity Research Group in Southwest Japan (1994) shows active seismic faults clearly as gravimetric lineament here and there. This figure is an example in the Tokai area , Japan. The most significant faults among them are Akaishi Fracture Zone along the Tenryu River and its branch eastward to Omaezaki Peninsula. By the helicopter surveys conducted in June 2002 we have obtained four profiles of gravity disturbances along east-to-west tracks (See Fig.6). The northernmost track crosses the Akaishi Fracture Zone and its branch on the landward area, and the other three tracks to the south with 5 nautical mile spacing cross the Enshu-Nada canyon which seems to be an extension of the Tenryu River.

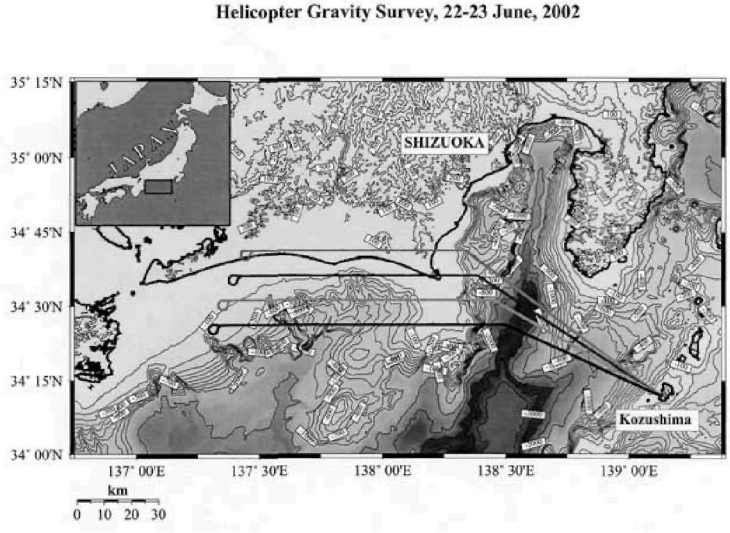


Fig.6 Four tracks of helicopter measurement conducted in June 2002. We used a heliport at Kozushima Island as a base station. The helicopter flied from Kozushima, crossed the Suruga trough and then took an east-to-west course to the Azumi Peninsula. The flight altitude is 2000 ft and the track spacing is 5 nautical miles.

measurements showing a beautiful agreement with land gravity data become disagreeable in turn at sea, there must be areal bias offset with the marine gravity data. According to Kuroishi et al. (2002) the undulation of geoid in the area concerned is less than 10 m in amplitude. So, there may be more or less 3 mgal ambiguity in the comparison made above between gravity disturbance and free air gravity anomaly, but the discrepancy as large as 15 mgal cannot be accounted for. It should be emphasized that this type of error has important effect on the estimation of geoid both on land and at sea.

5. Retracing active seismic faults from land to sea floor (mission 2)

Figure 5 shows Bouguer gravity anomaly with contour intervals of 0.5 mgal in the Tokai area of Honshu, Japan, together with four profiles of gravity disturbance measured by helicopter in the Enshu-Nada Sea at the altitude 2000 ft. The profiles by helicopter are obtained along the east-to-west parts of the tracks between 137° 30′ E and 138° 20′ E as shown in Figure 6. The northernmost track passes land area. The profile of gravity disturbance on each track is approximately 80 km long and runs with a spacing of 5 nautical miles (9.3 km). The profiles from the upper to the lower correspond to the tracks from north to south in order.

From the map of Figure 5 it is seen that the helicopter measurements cross two significant faults trending north-to-south: One is the Akaishi fracture zone running parallel to the Tenryu River (left north-to-south arrow in Figure 5) and the other is a branch of the Akaishi fracture zone (right north-to-south arrow in Figure 5). These two faults indicated by Bouguer gravity anomaly are also clearly reflected on gravity disturbance measured by helicopter along the northernmost track. The gravity anomaly profiles shown in Figure 5 consist of three lines each, i.e., outgoing and return gravity disturbances by helicopter and the sea truth free air gravity anomalies from marine measurements. It is seen that the helicopter gravity disturbances and the marine free air gravity anomalies appear harmonious from the rough sense of the word. If the four gravity profiles measured by helicopter are compared to each other, it is seen that the gravity gradient coinciding with the active faults on land is exaggerated seaward. It seems that, at first sight, the phenomenon reveals the development of faults beneath the sea floor, but it is unlikely; the relevant zone is an oceanic trough zone which causes negative gravity anomaly by topographic depression, making it uncertain whether the gravity gradient can be directly attributed to the sub-bottom fault or not.

6. Summary

A new model of helicopter-mounted gravimeter was developed in 1998 and has been used to measure gravity at the land-to-sea boundary zones around the Japanese Islands. Among the areas so far surveyed we report the measurements at Saitama, Ibaraki and Kashima-Nada Sea and the measurements in the Tokai District and the Enshu-Nada Sea. From these two measurements we have attained two findings as explained below

1) Regional discrepancy of gravity data in the Kashima-Nada Sea:

East-west measurements by helicopter-mounted gravimeter across the coast from Saitama, Ibaraki offshore to the Kashima-Nada Sea shows discrepancies between sea gravity measurements and helicopter measurements clearly, indicating that the gravity of marine measurements includes positive errors amounting roughly to 15 mgal. This finding would assuredly affect the geoid evaluation in the future.

.

2) Challenge to identify sea floor active faults:

It is known that accurate land gravity data delineate the trend of active faults, and it seems natural to consider that the faults stretching over to the coast is continued to the sea floor. In the case of Enshu-Nada Basin the Bouguer gravity anomaly clearly indicating the hidden active faults on land is disturbed at sea by the topography of the sea floor, i.e., the Enshu-Nada Canyon. It is true that the significant slope of gravity anomaly associated with the faults is exaggerated toward the sea, but it is difficult to judge if the active faults and the side slope of canyon are essentially the same or not.

Acknowledgements: The authors would like to express their sincere thanks to NEDO (New Energy and Industrial Technology Development Organization, Japan) for the financial support to develop the airborne gravimetry. They also appreciate the financial aids from the Institute of Oceanic Research and Development, Tokai University and the Earthquake Research Institute, University of Tokyo.

References:

Gravity Research Group in Southwest Japan (1994): A Bouguer gravity anomaly map in Central Japan, appended map of report, Geol. Surv. Japan, 280, 29-36

Kuroishi, Y., H. Ando and Y. Fukuda (2002): A new hybrid geoid model for Japan, GSIGEO2000, J. Geodesy, 76, 428-436

Peters, M. F. and J. M. Brozena (1995): Methods to improve existing shipboard gravimeters for airborne gravimetry, Proc. IAG Symposium G4, IUGG 1995, Boulder, Colorado July 2-14, 1995

Segawa, J., H. Hasegawa, T. Ishihara, S. Sakuma and M. Ishii (2000): A preliminary study of the development of a helicopter-mounted gravity measuring system, Bull. Inst. Oceanic Res. & Develop., Tokai Univ., 21,1-10

Segawa, J., S. Kusumoto, T. Ishihara, E. John Joseph, M. Komazawa, H. Hasegawa, S. Sekizaki, K. Takada, H. Suzuki, E. Nakayama, and S. Sakuma (2001): Results of gravity measurements (1999) using a helicopter-mounted gravimeter, Bull. Inst. Oceanic Res. & Develop., Tokai Univ., 22, 1-14

Segawa, J., E. J. Joseph, S. Kusumoto, T. Ishihara, E. Nakayama, and M. Satomura (2002): Development of the helicopter-mounted gravimeter and the study of the active faults running across the coastal lines over the continental shelves, Proc. 3rd Meeting of the International Gravity and Geoid Commission (IGGC), I. N. Tziavos (Ed.), Gravity and Geoid 2002-GG2002, Thessaloniki, Greece, August 26-30, 2002.

A Gravity Database of Southwest Japan: Application to Bouguer Gravity Imaging in Kyushu District, Southwest Japan

R. Shichi

College of Engineering, Chubu University, Kasugai, Aichi 487-8501, Japan

A. Yamamoto

Graduate School of Science, Hokkaido University, N10S8, Kita-ku, Sapporo 060-0810, Japan

T. Kudo

Tono Geoscience Center, Japan Nuclear Cycle Development Institute, 1-63, Yamanouchi, Akeyo, Mizunami 509-6132, Japan

Y. Murata, K. Nawa, M. Komazawa

The Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology, Tsukuba Central 7, 1-1, Higashi 1-Chome Tsukuba, Ibaraki 305-8567, Japan

M. Nakada

Graduate School of Science, Kyshu University, 6-10-1 Hakozaki, Fukuoka 812-8581, Japan

H. Miyamachi

Faculty of Science, Kagoshima University, 1-21-35 Korimoto, Kagoshima 890-0065, Japan

H. Komuro

Interdisciplinary Faculty of Science and Engineering, Shimane University, 1060 Nishikawatsu, Matsue 690-8504, Japan

Y. Fukuda, T. Higashi, Y. Yusa

Graduate School of Science, Kyoto University, Kyoto 606-8502, JAPAN

I. Nakagawa

Daiichi Fukken Engineering Co., Ltd., 3-5-28 Fukuokaeki-minami, Fukuoka 812-0016, Japan

H. Watanabe, J. Oikawa

Earthquake Research Institute, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

S. Kobayashi

Space and Earth Information Technology, School of Engineering, Kyushu Tokai University, Kumamoto 865-8652, Japan

I. Ohno

Department of Earth Sciences, Faculty of Science, Ehime University, 10-13, Dogo-Himata, Matsuyama 790-8577, Japan

Abstract. In order to investigate subsurface density structures with the highest precision, two major gravity databases and some unpublished data files, which cover Southwest Japan, are recompiled. This paper presents some examples of the detailed gravity anomaly maps in Kyushu District, southwest Japan. These maps indicate various fine subsurface structures, which may arose with tectonic movements and/or volcanic activities of this region. With an aid of geological and topographical information, we discussed relationships between characteristic patterns of the detailed gravity anomalies and subsurface structures in this region. The most conspicuous feature is an existence of a steep horizontal gradient zone of gravity anomalies running through the Aso Caldera in the ENE-WSW direction. Although the location of the westward extension of the Median Tectonic Line (MTL) in this region is still debatable, this zone seems to be the westernmost part of MTL.

Keywords. Gravity database, Gravity anomaly, Kyushu district, Southwest Japan

1 Introduction

Two large gravity databases in Japan have been successively issued for public use in 2000 and 2001. One is the "Gravity CD-ROM of Japan" (GSJDB) published by the Geological Survey of Japan (GSJ) (GSJ, 2000). The other is the "Gravity Database of Southwest Japan" (SWJDB) published by the Gravity Research Group in Southwest Japan (GRGSWJ) (2001). Total number of original gravity data published by these two databases amounts to 136,698 and 90,298 points, respectively. Data coverage of the GSJDB extends to whole Japanese Islands, while the SWJDB covers largely southwest Japan. Both databases were formed mutually complementary in their distributions of gravity data. Consequently, by combining these two databases, gravity data of southwestern area of the Japanese Islands lose their non-closely-spaced character and spread evenly over the area. By adding unpublished gravity data provided by the Geographical Survey Institute (GSIDB), we compiled a gravity database in southwest Japan. Thus, southwest Japan is successfully covered by uniformly- distributed dense gravity data with no blank area of observation left.

2 Recompiling of the gravity databases

Although the databases above have been compiled

with a higher precision of about 1 mgal, they still contain gravity data with a precision no better than a few milligals or more which are not accurate enough for a detailed mapping. We (members of GRGSWJ) are now executing further revisions to remove those poorly-calibrated data by careful scrutiny. We have accomplished this revision in the area of Kyushu District, southwest Japan, and it becomes possible to draw very precise gravity anomaly atlas which fully guarantees a precision of 1 mgal in the whole area.

3 Bouguer gravity imaging in Kyushu District, Southwest Japan

In constructing Bouguer anomaly atlas, we used several reduction density values, ranging from 2.2 to 2.67 g/cm^3, to remove the correlation between gravity anomaly and surface topography particularly in volcanic areas. Examples of the results obtained through this procedure are shown in Figs. 1 and 2. Fig. 1 is a Bouguer anomaly map contoured at 2 mgal interval for a density 2.67 g/cm^3 and superposed on the active faults in Kyushu District. Fig. 2 is one for a density 2.3 g/cm^3 contoured by the same manner as Fig.1. Gravity features of the whole Kyushu District are delineated very precisely on those figures. It is reasonable to adopt a density value of 2.67 g/cm^3 for an overview of the whole Kyushu area, while a density value of around 2.3 g/cm^3 is best fit for the local area of volcanic zones. A low density yields artificial complicated patterns in the mountainous area other than volcanic zones (cf. Kyushu Mountains area in Fig.2).

4 Characteristic features shown in the detailed gravity anomaly maps of Kyushu District

The Median Tectonic Line (MTL) is extending more than 800 km in the median zone of Southwest Japan. The northern Kyushu, in the geologic sense, have characteristic of the northern side (Inner Zone) of the MTL, and southern Kyushu have that of the southern side (Outer zone) (Huzita, 1980). Nevertheless, only the central Kyushu is characterized by many normal faults, huge active volcanoes, and active seismicity in the crust (e.g., Kakuta et al., 1992). Characteristic features found from the new precise gravity anomalies in Kyushu District are summarized as follows;

(1) The Aso volcano, located in the central Kyushu, has one of the largest craters (caldera) in the world. Yokoyama (1963) initially pointed out that the Aso Caldera is characterized by a prominent low gravity anomaly. Further gravity investigations made sure that this gravity depression is nearly circular shaped, with a steep rim and a flat bottom accompanying several closed minima (e.g., Komazawa, 1995). The new Bouguer anomaly map shows that gravity changes along the rim have the amplitude of more than 10 mgal/km. This feature implies that the Aso Caldera has not the funnel-shaped structure, which is suggested by Yokoyama (1963), but a piston-shaped cylinder structure.

(2) A dominant steep gravity gradient zone which is running through the Aso Caldera is observed along the ENE-WSW trending The Oita-Kumamoto Tectonic Line (OKTL). OKTL is running nearly along the Tertiary-PreTertiary geologic boundary in central Kyushu. The eastward extension of this gradient zone corresponds to the location of the MTL at Shikoku District. Therefore, it is strongly

suggested that the westward extension of the MTL is along the gradient zone. On the other hand, southwestward continuation of this gradient zone is passing through the northern coast of the Uto Peninsula and can be traced to Hondo (central Amakusa, southern part of the Inner Zone of Kyushu). Adjacent to the north of this steep gradient zone lies a huge and strong low anomaly belt, which is passing through Iyonada - Beppu Bay - Kuju Volcano - Kumamoto - Unzen Volcano - Tachibana Bay - Amakusanada from east to west.

(3) A strong negative Bouguer anomaly is observed over the Hyuganada Basin, southeastern Kyushu, and characterized by a sharp inland ward increase of Bouguer anomaly. This gravity low forms arcuate depression whose northwestern boundary is sharply cut by Pre-Tertiary volcanic rocks. This implies that subsurface structure beneath the basin may be correlated to the tectonic drug caused by subduction of the Philippine-Sea Plate.

(4) Trend-reduced Bouguer anomaly shows a large positive anomaly over the Pre-Tertiary formations in the northern part of Kyushu. In contrast, negative anomalies are dominant over the Pre-Tertiary formations in the southern part of Kyushu, particularly around the Kyushu Mountains. While the Amakusa Islands in western central Kyushu, probably the westernmost part of MTL, are characterized by a strong gravity high in detrended Bouguer map, which is well corresponded with Paleogene formations.

(5) Locations of active faults (Nakada and Imaizumi, 2002) are also depicted in Fig. 1. It is obvious that many parts of active faults correspond with narrow gravity cliff or ridge. Moreover, some of these gravity lineaments are much longer than the corresponding active faults. These parts of lineament presumably indicate subsurface extensions of the fault structures.

5 Conclusions

Detailed gravity anomaly maps in Kyushu District are presented. Characteristic patterns and geometries in the gravity anomaly distributions in this region provided much helpful information for exploring hidden faults and other subsurface geologic structures. Especially, Bouguer anomaly maps derived from the revised database might be some useful constraints for understanding of the tectonics at the western end of the MTL.

Acknowledgments

We would like to express our thanks to the Geographical Survey Institute, Japan and Nihon Chinetsu Kaihatsu Sokushin Center, Japan for the permission to use the gravity data.

References

The Geological Survey of Japan (ed.) (2000), Gravity CD-ROM of Japan, Digital Geoscience Map P-2, Geological Survey of Japan.

The Gravity Research Group in Southwest Japan (Representatives: Ryuichi Shichi and Akihiko Yamamoto) (2001), Gravity Measurements and Database in Southwest Japan, Gravity Database of Southwest Japan (CD-ROM), Bull. Nagoya University Museum, Special Rept., No.9.

Huzita, K. (1980). Role of the Median Tectonic Line in the Quaternary Tectonics if the Japanese Islands. Mem. Geological Soc. Japan, 18, pp. 129-153.

Kakuta, T., H. Miyamachi, H. Kubota and A. Takagi (1992). Crustal Earthquakes in Kyushu. Zisin. Journal of the Seismological Society of Japan, 45, pp. 229-237.

Komazawa, M. (1995). Gravimetric Analysis of Aso Volcano and its Interpretation. J. Geod. Soc. Japan, 41, 17-46.

Nakada, T. and T. Imaizumi (ed.) (2002). Digital Active Fault Map of Japan. Univ. of Tokyo Press.

Yokoyama, I. (1965). Structure of caldera and gravity anomaly. Bull. Volcanol., 26, pp. 67-72.

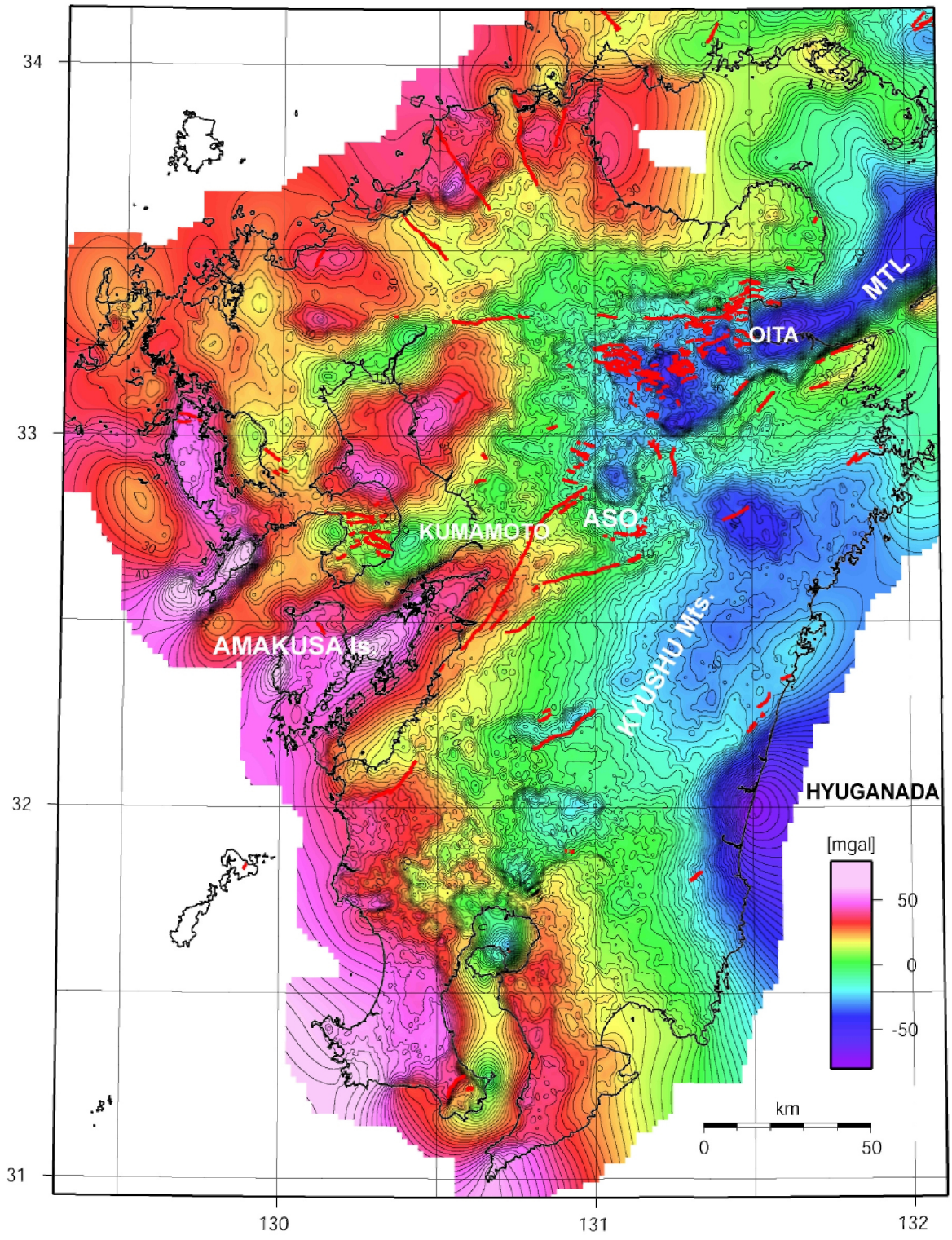


Fig. 1. Bouguer anomaly map of Kyushu District. Assumed density: 2.67. Contour interval: 2 mgal. Red line segments: Active faults.

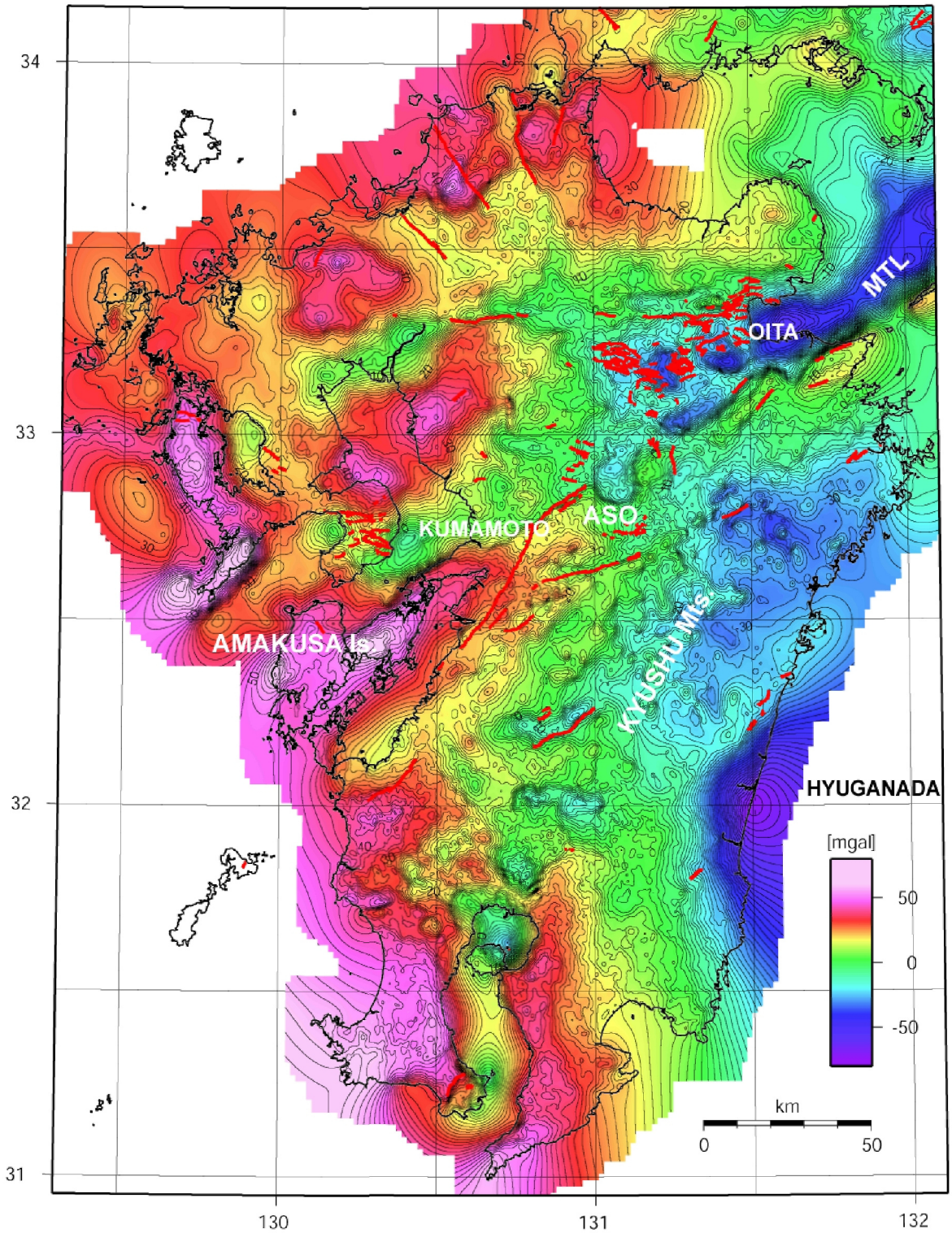


Fig. 2. Bouguer anomaly map of Kyushu District. Assumed density: 2.30. Contour interval: 2 mgal. Red line segments: Active faults.

Gravity-Based Fault Mapping: The Ishikari Lowland, Hokkaido, Japan

A. Yamamoto

Institute of Seismology and Volcanology, Graduate School of Science, Hokkaido University, N10–W8, Kita–Ku, Sapporo, 060–0810, Japan

Abstract. This paper examines gravity structures of the Ishikari Lowland of Hokkaido, Japan, by focusing relief-shaded Bouguer gravity, specifically to relocate the faults (the Ishikari Teichi Touen Fault Zone, ITTFZ), and to present gravity-based fault mapping as another approach for estimate of fault segmentation. Gravity analyses are strictly based on dense gravity data measured by various institutes and organizations. Bouguer relief is produced by illuminating the light from eight directions to effectively display the detailed gravity features varying laterally along the azimuthal direction. A striking linearity of the relief-shaded Bouguer gravity along the ITTFZ is found on most of the relief maps, particularly on the maps for the azimuth of the due east and west directions. In the central part of the ITTFZ, however, the lineament in the relief-shaded Bouguer map does not bear a good correlation with the known fault distributions. In addition, the gravity relief for the azimuth of the south-west direction exhibits a remarkable lineament, extending southward from near Bibai, whose southward continuation can be traced to the south-east until 20~25 km south of Atsuma along the westernmost boundaries of pre-Neogene volcanics. This implies that the geometry of the southern end of the ITTFZ provides distinct continuity along the relief-shaded Bouguer lineament roughly to the south-east.

Keywords. Gravity anomaly, Fault mapping, Ishikari Plain, Ishikari Lowland, Hokkaido, Sapporo, Japan

1 Introduction

The Ishikari Plain is chief industrial region and characterized by the largest alluvial lowland (the Ishikari Lowland) in Hokkaido, Japan. Fig.1 shows a 3-D perspective view of topography with known faults (Fig.1a), and a simplified geologic map (Fig.1b) around the Ishikari Lowland, where Bouguer anomaly isolines are superimposed. Geology information in Fig.1b is based on the digital version of 1:1,000,000 scale geological map of Japan (3rd edition) by Geological Survey of Japan (1995). Thick colored lines and dotted red lines in Fig.1b denote known active faults by Nakata and Imaizumi (2002) and the Research Group for Active Faults of Japan (1991), respectively. Thick color-coded faults in Fig.1 indicate, red: fault certainly exists and its location is accurately determined, magenta: fault certainly exists but its location is not accurate, green: fault possibly exists but invisible, and blue: estimated fault lying at depths, respectively. As shown in Fig.1, the Ishikari Plain is characterized by the largest late-Cenozoic lowland (the Ishikari Lowland) in west Hokkaido. Wide-spread deposition of alluvial deposits constitutes a large part of the overall sedimentary sequence around the plain. Fault-bounded plains or basins often accompany a variety of tectonic motions such as reverse faulting. Specifically in Hokkaido, these motions are associated with the plate movements and the secondary tectonics. In the eastern margin of this lowland lies the Ishikari Teichi Touen Fault Zone (ITTFZ) which borders on the Miocene hill belts (Iwamizawa, Kurisawa and Umaoi hills). The ITTFZ has en echelon features and runs nearly in the N-S direction, bending toward west, and is classified as reverse fault system. (Research Group for Active Faults of Japan, 1991; Oka, 1986; Oka et al., 2001). The gravity field prevailing over the Ishikari Plain is best reflected on the Bouguer anomaly map. Recently, Yamamoto (2003a,b) gave preliminary results of Bouguer anomalies over the plain and discussed about their important features. This paper examines gravity structures of the Ishikari Lowland by focusing Bouguer gravity reliefs, specifically (1) to relocate the faults (the ITTFZ), comparing the

new active fault distributions (Nakata and Imaizumi, 2002) with the older ones (Research Group for Active Faults of Japan, 1991), and (2) to present the gravity-based fault mapping as another approach for correct estimate of fault segmentation.

2 Gravity Data

In total, about 25,000 gravity data from Hokkaido University, Geological Survey of Hokkaido (hereafter referred to as GSH), the Japan Petroleum Exploration, Co. Ltd. (hereafter referred to as JAPEX), etc. were collected for gravity study around the plain. Particularly, gravity data measured by JAPEX number more than 20,000 in the lowland and its vicinity with an average spacing of 300~400 m. Komazawa et al. (1998) published a Bouguer anomaly map around the Ishikari Plain assuming a reduction density of 2.3 g/cm^3 on the basis of land gravity data by JAPEX. In the present study, gravity values for land stations obtained by JAPEX were first recalculated to determine absolute values based on the Japan Gravity Standardization Net 1975 (JGSN75, Suzuki, 1976). Then, JGSN75-based values were converted to the gravity values based on the Japan Gravity Standardization Net 1996 (JGSN96, Nakai et al., 1997; Yamaguchi et al., 1997). Land gravity data by GSH were compiled and also recalculated, referring to the values of JGSN96, so as to fit the reference gravity data by Hokkaido University to which the absolute gravity values were already assigned. Gravity data by Hokkaido University were obtained by a LaCoste & Romberg Model G land gravity meter and a Scintrex gravity meter. These data were fully corrected for latitudinal and elevation effects, instrumental height, earth tide, and secular drift of the spring. These gravity data were then reduced for free-air and Bouguer corrections, as well as for terrain effects, where spherical terrain corrections were applied to a radius of 80 km according to the method by Yamamoto (2002). In all reductions the earth's sphericity is taken into consideration. After an overall revision and reduction for these gravity data, a Bouguer anomaly map, a residual anomaly map, and a first horizontal derivative map of the region were constructed.

3 Relief-Shaded Bouguer Gravity

As shown in Fig.1b, gravity field over the Ishikari Plain is characterized by distinct low anomaly as compared to its surroundings. This gravity low is mostly associated with significant Quaternary sediments spreading extensively over the area. Steep gravity gradients are observed along the ITTFZ in the eastern margin of the plain. They are trending almost in the N-S direction, which is closely correlated with that of the active fault distributions. While gravity anomalies around the Nopporo Hill (NH), where the Nopporo Hill Fault Zone (NHFZ in Fig.1a) is newly identified as an active fault system in the newly-compiled fault mapping (Nakata and Imaizumi, 2002), are small in amplitude and are associated with no abrupt gravity changes compared to its topography and surrounding fault distributions, implying that no sharp density contrast are appreciable at depths. To clearly show the gravity structure around the Ishikari Lowland, particularly the sharp and fine linearity corresponding well to the ITTFZ, we discuss here relief-shaded gravity anomaly illuminated by the light from various directions. In general, Bouguer anomaly can be shaded by giving illumination to highlight gravity structures laterally varying along the azimuthal direction (the direction of illumination). Therefore, gravity structures with lateral variations along the azimuthal direction display striking contrast and also show striking difference as compared with those with lateral variations perpendicular to the azimuthal direction. This means that gravity structures with lateral variations perfectly perpendicular to the azimuthal direction never appear in gravity relief map shown in shades of white and black. We first take a horizontal derivative of non-filtered Bouguer anomaly along the azimuthal direction. As is well known, horizontal gradients of gravity anomalies have a multiresolution interpretation based on Poisson kernel (Green's function for potential fields). Then we normalize the derivative such that their values range from −1 (black) to 1 (white). This procedure makes brighter the areas whose Bouguer values increase along the illumination vector, while the areas with decreasing values along the vector are shown gloomily (blackish color in the figure). Note that we simply change the azimuth of view point and use the fixed elevation of zero degree (horizontal view) in this study. In this way relief-shaded Bouguer gravity maps were produced for eight directions of illumination. Fig.2a illustrates the relief-shaded Bouguer gravity map in a planimetric view illuminated by the light from the south-west direction. This is quite effective to make clear the laterally heterogeneous gravity structure varying along SW-NE direction around the ITTFZ. Note that the illumination vector is shown on the upper-right side of the figure. Fig.2b is the same as Fig.2a, but known active faults are superimposed in

bright colors. Heavy colored lines and dotted red lines in Fig.2b depict known active faults by Nakata and Imaizumi (2002) and the Research Group for Active Faults of Japan(1991), respectively. Colored faults in Fig.2b indicate, red: fault certainly exists and location is accurately determined, magenta: fault certainly exists but location is not accurate, green: fault possibly exists but invisible, and blue: estimated fault lying at depths, respectively. Characteristic features in Fig.2 are summarized below.

(a) A pronounced linearity of shaded relief along the ITTFZ is uncovered. This lineament extends from at least 5 km north of IWZ (Iwamizawa) to several kilometers west portion of HKT (Hayakita) and ATM (Atsuma), and also shows a good agreement with distributions of the old (dotted red line) and new (thick colored line) active faults in northern and southern parts of the ITTFZ (Fig.1b). Whereas, in central part near NNM (Naganuma), the lineament of relief-shaded gravity does not agree with the trend of active faults.

(b) Two high gravity belts, running almost in the N-S direction from near Iwamizawa (IWZ) to near Hayakita (HKT), can be found distinctly. These striking two belts are depicted as a pair of white and black shades which is protruding westward. Note that each of these two belts consists of a pair of Bouguer high and low extremes. Both of these two belts develop and are mapped in the western margin of the Miocene hill belts.

(c) In contrast, gravity low is clearly observed along the new active fault extending northward from near KYM (Kuriyama) to near BBI (Bibai). This feature can be stressed particularly on the gravity relief map for the azimuthal direction of 0° (due north) or 180° (due south) from north clockwise. This low anomaly runs along the eastern margin of the Iwamizawa Hill (IH) and the Kuriyama Hill (KH). However, it does not continue southward from KYM, implying that a sharp gravity change does not extend toward south along the eastern margin of the Umaoi Hill (UHN, UHC, UHS).

(d) The southernmost extension of the lineament described in (a) reaches about 5 km south of ATM (Atsuma) with good and unbroken continuity. This fact suggests that active fault zone along the ITTFZ is spatially distributed over a distance of more than 80 km, whose southward extension can be mapped until about 20 km south of ATM along the boundaries of pre-Neogene volcanic rocks.

The fact (d) is the most intriguing result in this study since the southward extension of the ITTFZ can be well defined and mapped by the present gravity relief analysis, where no noteworthy faults are distributed. From this result we infer that the geometry of the southern end of the known fault system (the ITTFZ) provides distinct continuity over a distance of 20~25 km along the relief-shaded Bouguer lineament roughly to the south-east. Consequently, our results suggest that the relief-shaded Bouguer gravity analysis, using closely spaced Bouguer data, is quite effective in demarcating the extension of the pre-existed (or invisible) faults for correct estimates of their location and segmentation.

4 Conclusions

Bouguer gravity relief was produced by illuminating the light from eight directions to effectively depict the detailed gravity features varying laterally along the azimuthal direction. A conspicuous linearity of shaded relief along the ITTFZ is detected from Bouguer anomaly relief particularly on the map for the azimuth of the south-west direction. This lineament extends from at least 5 km north of IWZ (Iwamizawa), ending over several kilometers west of HKT (Hayakita) and ATM (Atsuma). Also, this lineament shows a good agreement with distributions of the old and new active faults in northern and southern parts of the ITTFZ. Furthermore, a remarkable lineament extending southward from near BBI to south of ATM is found on gravity relief map particularly for the azimuthal direction of 0° (due north) or 180° (due south) from north clockwise. This fact asserts that the active fault strands along the ITTFZ is spatially distributed over a distance of more than 80 km and its southward extension can be mapped until 20~25 km south of ATM, separating the pre-Neogene volcanic rocks. Concludingly, we note that relief-shaded Bouguer gravity analysis plays an important major role in the gravity-based fault mapping to locate and characterize the various strands and segmentation of a fault zone.

Acknowledgments. The author gratefully acknowledges the personnel at the Japan Petroleum Exploration, Co. Ltd., who kindly permitted to use their land gravity data. This paper benefited from constructive criticism made by J. A. R. Blais. All figures were produced using the GMT (Generic Mapping Tools) software of Wessel and Smith (1998).

References

Geological Survey of Japan (ed.) (1995). Geological Map of Japan 1,000,000, 3rd Edition, CD-ROM Version, Digital Geoscience Map G-1 (DGM-G-1), Geological Survey of Japan.

Komazawa, M., T. Hiroshima, Y. Murata, M. Makino and R. Morijiri (1998). A gravity anomaly map of Sapporo, 1:200,000, Gravity Map Series, 10, Geological Survey of Japan.

Nakai, S., K. Yamaguchi, K. Nitta, H. Yamamoto, K. Matsuo, M. Machida, M. Murakami, M. Ishihara, R. Shichi and A. Yamamoto (1997). Data processing for the Japan Gravity Standardization Net 1996, in "Gravity, Geoid and Marine Geodesy" (GraGeoMar96), Proceedings of the International Symposium, No.117, Tokyo, Japan, September 30 - October 5, 1996, convened and edited by J. Segawa, H. Fujimoto and S. Okubo, pp.228–233, Springer-Verlag Berlin Heidelberg, (pp.746), ISBN:3-540-63352-9.

Nakata, T. and T. Imaizumi (ed.) (2002). Digital active fault map of Japan, University of Tokyo Press, pp60, (in Japanese).

Oka, T. (1986). Distribution and tectonic evolution of Late Cenozoic basins in Hokkaido, *Monograph Assoc. Geol. Collab. Japan*, 31, pp.295–320, (in Japanese).

Oka, T., J. Tajika, N. Ohtsu, W. Hirose, N. Okazaki and S. Ishimaru (2001). The Ishikari Teichi Touen Fault Zone, Active fault map and explanations, Geological Survey of Hokkaido (ed.), Hokkaido Government, pp157, (in Japanese).

Research Group for Active Faults of Japan (1991). Active faults in Japan, sheet maps and inventories, revised edition, University of Tokyo Press, pp437, (in Japanese).

Suzuki, H. (1976). The International Gravity Standardization Net 1971 and the Japan Gravity Standardization Net 1975, *J. Geod. Soc. Japan, 22*, pp.112–129.

Wessel, P. and W. H. F. Smith (1998). New, improved version of the generic mapping tools released. EOS Trans. AGU 79, p. 579.

Yamaguchi, K., K. Nitta, H. Yamamoto, K. Matsuo, M. Machida, M. Murakami, M. Ishihara, S. Nakai, R. Shichi and A. Yamamoto (1997). The establishment of the Japan Gravity Standardization Net 1996, in "Gravity, Geoid and Marine Geodesy" (GraGeoMar96), Proceedings of the International Symposium, No.117, Tokyo, Japan, September 30 - October 5, 1996, convened and edited by J. Segawa, H. Fujimoto and S. Okubo, pp.241–248, Springer-Verlag Berlin Heidelberg, (pp.746), ISBN:3-540-63352-9.

Yamamoto, A. (2002). Spherical terrain corrections for gravity anomaly using a digital elevation model gridded with nodes at every 50 m, *J. Fac. Sci., Hokkaido Univ.*, 11, No.6, pp.845–880.

Yamamoto, A. (2003a). Gravity anomaly atlas of the Ishikari Plain and its vicinity, Hokkaido, Japan, *Geophys. Bull. Hokkaido Univ.*, 66, pp.33–62 (in Japanese).

Yamamoto, A. (2003b). Gravity-based active fault mapping around the eastern margin of the Ishikari Lowland, Hokkaido, Japan, *J. Fac. Sci., Hokkaido Univ.*, 12, No.1, pp.17–39, 2003.

A. Yamamoto, Institute of Seismology and Volcanology, Hokkaido University, Sapporo, 060–0810, Japan. (Email:star@eos.hokudai.ac.jp)

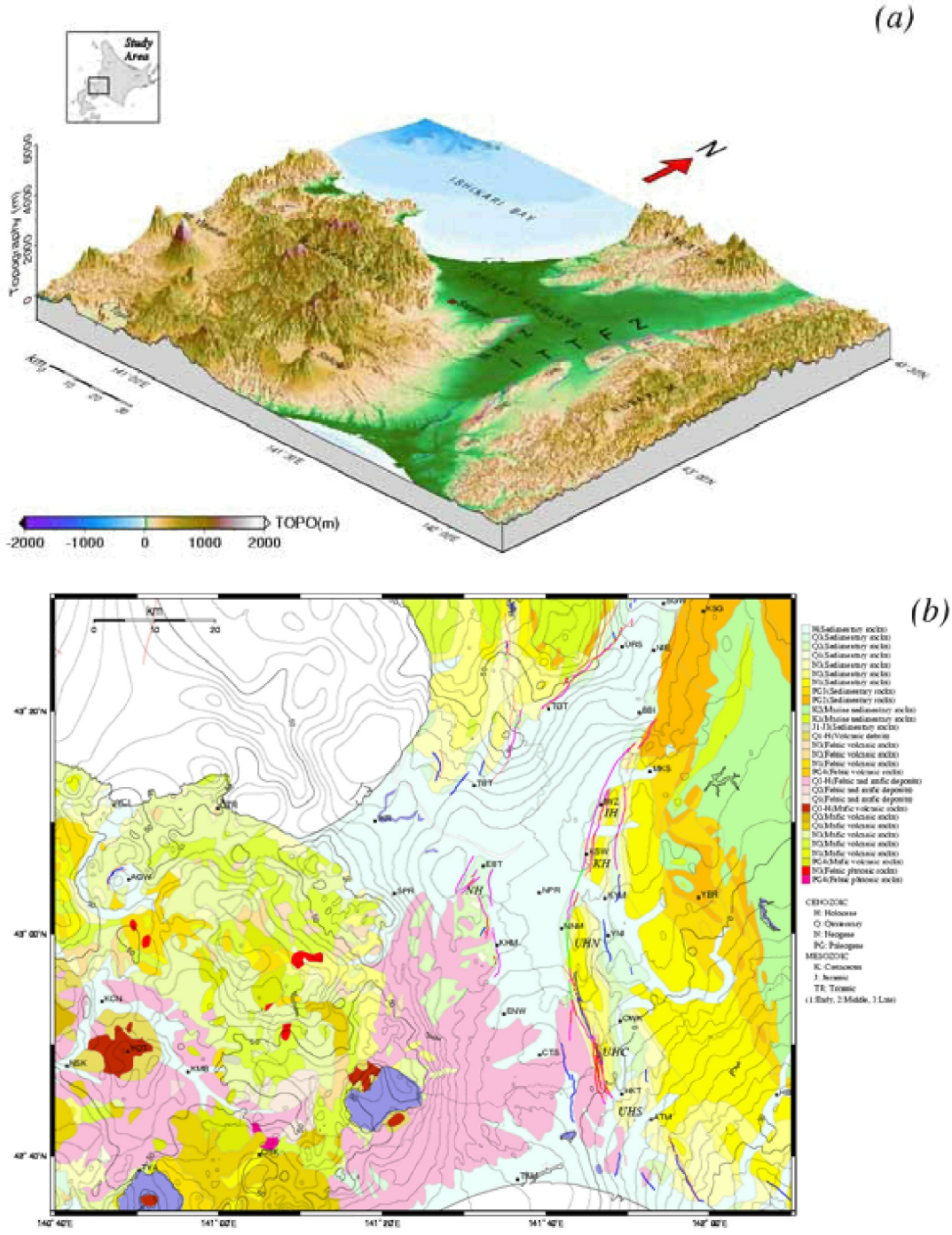


Fig. 1. (a) A 3-D perspective map showing topography and known faults around the Ishikari Lowland. Colored active faults are taken from Nakata and Imaizumi (2002). Vertical exaggeration is about 5 times. ITTFZ: the Ishikari Teichi Touen Fault Zone, IH: Iwamizawa Hill, KH: Kurisawa Hill, UHN, UHC, UHS: Umaoi Hill (Northern, Central and Southern Block), NHFZ: the Nopporo Hill Fault Zone. (b) Simplified geologic map around the Ishikari Lowland. Bouguer anomaly isolines are superimposed with a contour interval of 5 mgal. Assumed density is 2.67 g/cm^3. Geology information is taken from Geological Survey of Japan (1995). Heavy colored lines and dotted red lines demonstrate known active faults by Nakata and Imaizumi (2002) and the Research Group for Active Faults of Japan (1991), respectively. Large closed triangles and squares, followed by three letters, show geographical locations of major named summits and cities (towns), respectively. NH: Nopporo Hill, SPR: Sapporo.

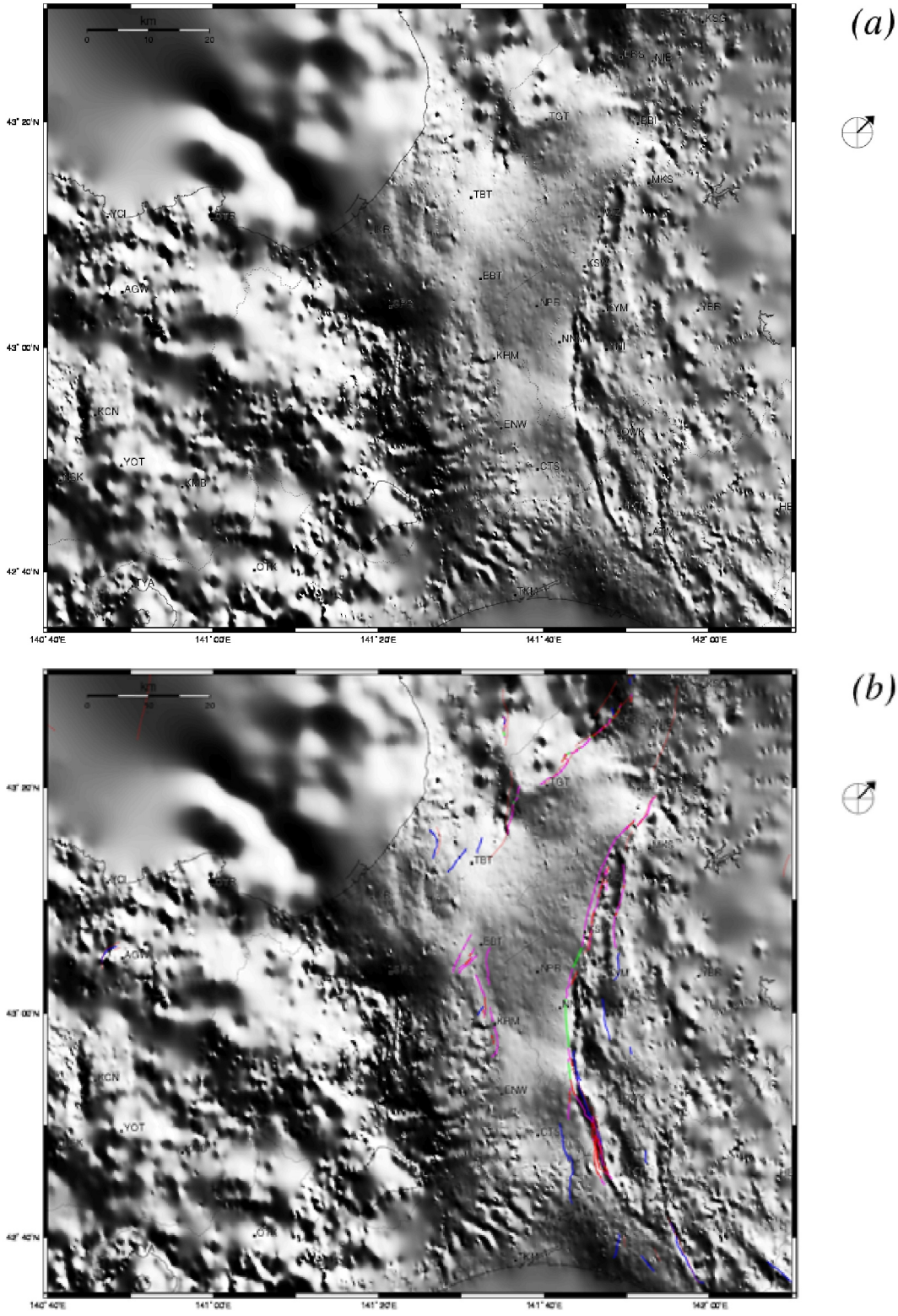


Fig. 2. (a) Relief-shaded Bouguer gravity map of the Ishikari Lowland and surrounding areas, illuminated by the light from the south-west direction (azimuth: N135°W, elevation: 0°). Azimuthal direction is shown on the upper-right side of the figure. No filtering is applied to Bouguer anomaly. (b) Same as (a), but known active faults are superimposed in bright colors. Heavy colored lines and dotted red lines denote known active faults by Nakata and Imaizumi (2002) and the Research Group for Active Faults of Japan(1991), respectively. Colored faults indicate, red: fault certainly exists and location is accurately determined, magenta: fault certainly exists but location is not accurate, green: fault possibly exists but invisible, and blue: estimated fault lying at depths, respectively.

Compilation and Evaluation of a Consistent Marine Gravity Data Set Surrounding Europe

Heiner Denker, Markus Roland
Institut für Erdmessung, Universität Hannover
Schneiderberg 50, D-30167 Hannover, Germany
Email: denker@ife.uni-hannover.de, Fax: +49 511 762 4006

Abstract. Various institutions have collected shipborne gravimetric measurements during the last decades. Due to different standards used for the processing of the observations and the necessary corrections, significant inconsistencies exist between different cruises. This contribution aims at producing a consistent marine gravity data set surrounding Europe, which can then be used for high precision geoid modelling, dynamic sea surface topography estimation, and other applications.

Besides our own marine gravity data holdings, data were collected from the Bureau Gravimétrique International (BGI), the National Imagery and Mapping Agency (NIMA, formerly DMA), and the National Geophysical Data Center (NGDC). The area of investigation is spanning the latitudes from 10 °N to 90 °N and the longitudes from 60 °W to 60 °E. The quality of the data varies between the individual cruises, as they originate from many projects at different epochs. Hence, systematic errors are likely to exist. Such errors can be significantly reduced by a crossover adjustment of the individual ship tracks. Because the track information was not available for all cruises, it had to be regenerated by different procedures. Furthermore, duplicate sources were removed before the crossover adjustment. The crossover adjustment is based on a bias per track error model. The adjustment of about 1.5 million observations in nearly 17,000 tracks led to a consistent high quality marine gravity data set. The RMS of the about 80,000 crossover differences is 15.5 mgal for the original data set, 8.4 mgal for an edited data set, and 4.7 mgal for the final crossover adjusted data set.

The second part of this contribution describes the evaluation of the marine gravity data set by altimeter derived gravity anomalies from different sources. These comparisons also prove the effectiveness of the crossover adjustment.

Keywords. crossover adjustment, shipborne gravity observations, altimetric gravity anomalies

1 Introduction

Shipborne gravimetric measurements, using sea gravity meters mounted on gyro stabilized platforms, have been performed around the world for several decades, see Torge (1989). Due to instrumental errors, navigational errors, and other error sources, significant inconsistencies exist between different cruises, cf. Wessel and Watts (1988). In the older marine gravity data sets, the major accuracy limiting factor was the ship's navigation, affecting the computation of the Eötvös effect. This problem could only be overcome with the GPS and other highly accurate positioning systems, but still exists in large parts of the marine gravity data.

The problem of inconsistencies in marine gravity data sets was treated successfully in the past by the analysis of intersecting ship tracks, e.g., Wessel and Watts (1988) analysed a global data set, while Strang van Hees (1983) and Wenzel (1992) studied regional data sets. From the crossover differences, error parameters per ship cruise or per track were estimated by a least-squares adjustment, yielding a significant reduction of the crossover differences. In this contribution, the marine gravity data, available for the European seas, are analysed using the crossover technique. In Section 2, the applied crossover analysis technique is outlined. Section 3 describes the data sources used in this study and the necessary preprocessing steps, e.g., the reconstruction of the ship tracks, and the removal of duplicate sources. The crossover adjustment results, based on a bias per track error model, are given in Section 4. Finally, an evaluation of the crossover adjusted data by altimetrically derived gravity anomalies is presented in Section 5.

2 Crossover Analysis Technique

Marine gravity observations are affected by instrumental errors (e.g., drift, cross-coupling, off-levelling, etc.), navigational errors, yielding incorrect Eötvös corrections and positions, and other error sources, such as incorrect ties to harbour

base stations and the inconsistent use of reference systems. In the older sea gravity observations, the major limitations are coming from inaccurate Eötvös corrections, resulting partly in errors of tens of mgals. This limitation could be overcome with the GPS and other highly accurate positioning systems, enabling now gravity accuracies below 1 mgal. Moreover, for some cruises, aiming mainly at local gravity investigations, the tie to harbour base stations was incomplete or not performed at all, and such cruises may be off by a few hundred mgal with respect to an absolute system.

For the majority of the cruises, it can be assumed that a correct tie to harbour base stations was done at the beginning and end of the cruise. Furthermore, from the harbour ties, drift corrections were usually derived and applied to the observations. In addition, in the newer observations, the drift problem has been further reduced by instrumental improvements. Therefore, a bias and eventually a tilt parameter per ship track is a reasonable error model, especially in view of possible errors in the Eötvös corrections, being constant for a ship track with constant heading and velocity-over-ground. In this connection, a ship track is considered as a part of the entire cruise, where the heading is almost constant. A tilt per track parameter, which could model, e.g., drift effects, is only appropriate for long tracks. From previous investigations it was found that a tilt parameter per track does not significantly improve the crossover differences at intersecting ship tracks. Therefore, in this study, a bias per track error model was used.

The basic principle of the crossover analysis technique is that the gravity value at intersecting ship tracks should by identical in both tracks involved. Thus for two tracks i and j and a bias per track error model, the gravity value at the crossover point, g_x, is

$$l_i + v_i + b_i = g_x, \quad (1)$$

$$l_j + v_j + b_j = g_x, \quad (2)$$

where l_i, l_j are the gravity observations of tracks i and j at the crossover point, v_i, v_j are the corresponding observation residuals for tracks i and j, and b_i, b_j are the bias parameters for tracks i and j. By combining the above equations, the crossover difference is obtained in the form of an observation equation for the adjustment:

$$(l_i - l_j) + (v_i - v_j) = d_{ij} + v_{ij} = b_j - b_i, \quad (3)$$

where $d_{ij} = l_i - l_j$ is the crossover difference, and v_{ij} is the corresponding residual.

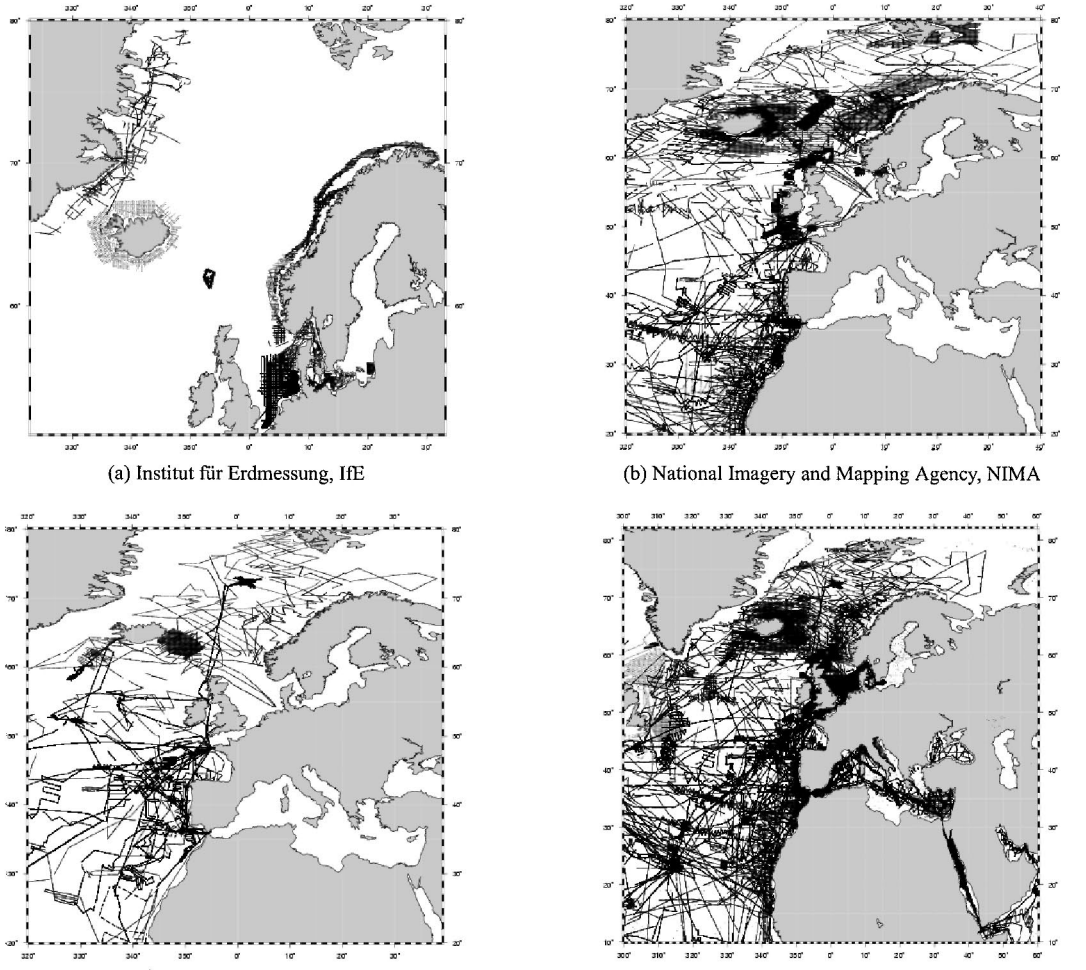
The above observation equations can now be used to estimate the unknown track biases by a standard least-squares adjustment procedure. The resulting normal equations can be solved under the assumption that all tracks have crossovers and that all tracks form a connected network, while the gravity datum has to be supplied either by constraining the track bias for one track to a given value or by a so-called free network adjustment, where the sum of all track biases is forced to zero. As there are often unconnected tracks, weak a priori informations are introduced for all track biases, allowing the solution of the normal equations in any case. This procedure is implemented in the program SEAGRA, being used in this study. The program was originally developed by Wenzel (1992), but the present version of the program includes several modifications regarding the crossover computation algorithm, the handling of unconnected tracks, and the output of statistical informations on a per track and per cruise basis. The program uses sparse matrix techniques and now can handle up to 50,000 tracks. The crossover points are detected automatically by a sophisticated iterative algorithm, which also works for curved tracks. The gravity values at the crossover points are computed by linear interpolation within the track, which was found to be sufficient, see also Wessel and Watts (1988). The adjustment can be done for one cruise only, giving the internal crossovers (within a cruise), or for many cruises, providing the internal and external crossovers (between different cruises).

3 Data Preprocessing

In this investigation, marine gravity data were selected for the area shown in Fig. 2. The data originate from the following four institutions (cf. Fig. 1):

- Institut für Erdmessung (IfE),
- National Imagery and Mapping Agency (NIMA, formerly DMA),
- National Geophysical Data Center (NGDC),
- Bureau Gravimétrique International (BGI).

For most of the individual ship cruises, unfortunately, the original track structure was not preserved. However, as this is the basic prerequisite for the application of the crossover analysis technique, this information had to be reconstructed. Basically three different procedures were used, depending on the original data formats and storage sequence, see also Behrend (1999). All procedures rely on the basic assumption that observations within one track have approximately the same azimuth and point separation. The first track reconstruction technique presupposes that the data are stored time-ordered in the original observation sequence. The segmen-



(a) Institut für Erdmessung, IfE

(b) National Imagery and Mapping Agency, NIMA

(c) National Geophysical Data Center, NGDC

(d) Bureau Gravimétrique International, BGI

Fig. 1. Marine gravity data collected from the four institutions IfE, NIMA, NGDC, and BGI.

tation of the entire cruise into tracks can then be accomplished quite easily by azimuth and distance criteria. The second segmentation procedure is used for cruises, which have given track numbers, but are not stored in the original observation sequence. This is, e.g., the case, if data were extracted from a data base ordered in geographical bins. In this case, data sorting within a track with respect to latitude, longitude, or distance yields the original track structure. The third case deals with data stored in arbitrary sequence and without any track numbers available. This most general and also most complicated case is handled iteratively using azimuth and distance criteria and often requires manual interactions, cf. Behrend (1999). For all three procedures, a low navigation accuracy, pretending a zigzag course, is a problem. However, in this investigation, the ship positions were not filtered and changed.

In a further preprocessing step, duplicate cruises were identified by computing the percentage of identical position and gravity values between all cruises. Thereby, duplicate cruises were found between the four different institutions (see above), but also within a single institution, where data sets were archived more than once. The duplicate sources were excluded from the following crossover analysis.

4 Crossover Adjustment Results

Altogether, 626 cruises with 2,491,697 observations without an error flag were collected from the four agencies IfE (24 cruises, 103,782 observations), NIMA (206 cruises, 458,626 observations), NGDC (56 cruises, 351,896 observations), and BGI (340 cruises, 1,577,393 observations). Within the pre-

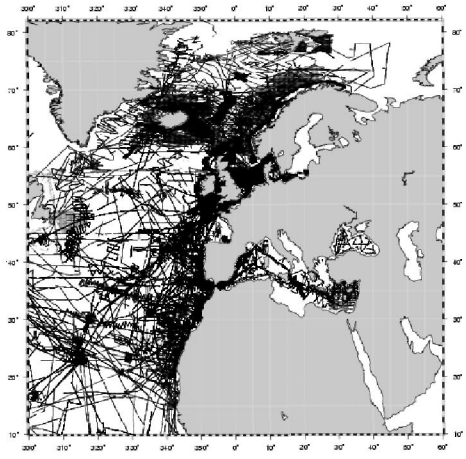


Fig. 2. Final edited gravity data set for the European seas.

processing step, 200 duplicate cruises with 697,644 observations were eliminated. Furthermore, additional cruises were excluded from the crossover adjustment because the reconstruction of the track informations was unsuccessful (36 sources, 49,796 observations), and because of extremely large (internal and external) crossover differences and discrepancies with altimetric data (5 sources, 24,829 observations).

Thus, the initial crossover adjustment of the entire data set was started with 385 sources and 1,719,428 observations. In the following, this data set is denoted as the "original data set". The crossover differences before the adjustment show a RMS value of 15.5 mgal and maximum values up to about 260 mgal, while after the adjustment the RMS value reduces to 7.0 mgal with maximum differences up to 205 mgal (cf. Table 1). The results from the crossover adjustment were analysed for gross errors, and a manual editing was done by inspecting the suspicious tracks. Often, problems were found at the beginning and end of the tracks, which is typically resulting from ship turns. Moreover, all data sources in the Red Sea area showed a very bad performance in the crossover adjustment and altimeter comparisons, and there were also problems in the regeneration of the track informations. Therefore, it was decided to exclude all data for the Red Sea, because these data sources would have required a very high effort for a careful data preprocessing and editing, which was considered as inadequate, especially because the Red Sea is outside our primary area of interest.

After the manual editing (42,039 observations) and exclusion of the Red Sea data (20 cruises, 149,033 observations), the crossover adjustment

Table 1. Statistics of crossover differences before and after adjustment for original and edited data set. Units are mgal.

data set	original		edited	
adjustment	before	after	before	after
#	89328	89328	78929	78929
mean	0.20	-0.02	0.04	-0.01
RMS	15.48	7.01	8.37	4.69
min	-258.43	-204.98	-109.91	-48.56
max	259.54	198.37	128.40	49.16

was repeated iteratively, and further editing of entire tracks was done. 64 short tracks with less than 3 points and 34 tracks with large crossover differences (mean>25 mgal, RMS>25 mgal, max.>50 mgal) were excluded (6037 points in total). In a final step, an editing of the unconnected tracks was done based on the comparison with the KMS02 altimetric anomalies (see Section 5). This led to the exclusion of 281 unconnected tracks with 13,800 points, where the differences exceeded the threshold values (mean>10 mgal, RMS>25 mgal, max.>50 mgal).

Thus, the final data set after the editing consists of 365 cruises with 1,508,519 points in 16,896 tracks, and will be denoted as the "edited data set" in the following. The locations of these data are displayed in Fig. 2. For the edited data set, 78,929 crossovers were detected by program SEAGRA (cf. Section 2). The statistics of the crossover differences before and after the adjustment are provided in Table 1, and corresponding histograms are given in Fig. 3. Table 1 clearly shows that the careful and time-consuming data editing is very important, resulting already in a reduction of the RMS crossover difference before the adjustment by a factor of two (15.5 mgal for the original data set vs. 8.4 mgal for the edited data set). For both, the original and the edited data set, the crossover adjustment reduces the RMS crossover difference by about a factor of two, leading to a RMS difference of 4.7 mgal for the final edited data set. Compared to the original data set, this is an improve-

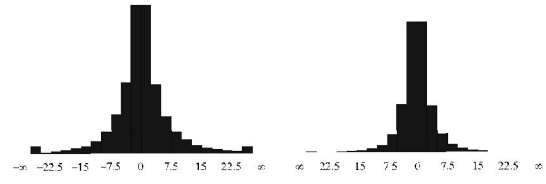


Fig. 3. Histograms of crossover differences before (left) and after (right) adjustment for the edited data set. Units are mgal.

Table 2. Statistics of crossover differences before and after adjusting the ship data separately by institution. Units are mgal.

	IfE		NIMA		NGDC		BGI	
	before	after	before	after	before	after	before	after
#	7091	7091	12481	12481	2655	2655	27035	27035
mean	0.16	-0.02	0.08	0.02	0.06	-0.07	-0.36	-0.04
RMS	4.41	2.35	5.41	3.71	8.20	3.82	10.23	5.00
min	-40.94	-22.84	-54.11	-37.47	-49.09	-28.97	-109.91	-46.72
max	51.55	25.66	60.88	45.19	45.45	28.34	128.40	49.15

ment by a factor of 3 to 4. Furthermore, Table 2 also provides the crossover statistics before and after adjusting the ship data (from the edited data set) separately by institution. The RMS differences vary between 4.4 mgal (IfE) and 10.2 mgal (BGI) before the adjustment, while the corresponding values after the adjustment range from 2.4 mgal (IfE) to 5.0 mgal (BGI). This shows that the level of accuracy differs for the four institutions, and once more the gain from the adjustment is about a factor of two for all institutions. In the end, it has to be noted that all adjustments were done with equally weighted crossover observations, as no reliable a priori error informations were available. This could eventually be improved in the future by deriving individual weights per cruise or per agency from the crossover analysis, cf. also Wessel and Watts (1988).

5 Evaluation by Altimetric Data

For the evaluation of the quality and consistency of the compiled marine gravity data set, altimetric gravity anomalies were utilized as an independent source of information. The altimetric results are available in grids and were interpolated to the locations of the ship observations by splines. Then differences of shipborne minus altimetric gravity anomalies were computed, and atmospheric corrections were considered for the shipborne data. The following four public domain altimetric data sets were used for the evaluation:

- CLS99, cf. CLS (2003),
- GSFC00, cf. GSFC (2003),
- KMS02, cf. Andersen and Knudsen (1998) and Andersen et al. (2003),
- Scripps 10, cf. Sandwell and Smith (1997).

Table 3. Statistics of the differences between the crossover adjusted shipborne and altimetric gravity anomalies. Units are mgal.

	CLS99	GSFC00	Scripps	KMS02
#	1399895	1400999	1400688	1400999
mean	-1.44	-1.04	-1.09	-0.97
RMS	9.21	8.18	8.20	7.78
min	-156.73	-158.47	-155.59	-157.73
max	115.91	102.02	106.46	106.00

The above altimetric gravity anomaly grids were computed directly from the sea surface heights (CLS99, GSFC00, KMS02) or via deflections of the vertical (Scripps 10), cf. the given references. The results of all comparisons with the final crossover adjusted marine gravity data set are summarized in Table 3. Unlike Fig. 2, the evaluation area was limited to 72 °N, because some of the altimetric grids do not extend further north. The mean difference is about -1 mgal for all four data sets, the reason for this being not clear at present. However, it should also be noted that the mean difference varies by geographical area, e.g., see Table 4. The RMS difference ranges from 9.2 mgal (CLS99) to 7.8 mgal (KMS02). Thus, the latest of the considered altimetric models, KMS02, shows the best agreement with the shipborne gravity data. The large minimum and and maximum differences are located in the vicinity of islands and in fjords, where the altimeter results may not be reliable.

In order to check the gain in accuracy due to the preprocessing and crossover adjustment, Table 4 provides the comparison results with KMS02 for the original and the edited data set. It becomes clear, that already the data editing significantly improves the agreement with the altimetry data. The

Table 4. Statistics of the differences between unadjusted and crossover adjusted shipborne and altimetric gravity anomalies (KMS02) including a subset. Units are mgal.

data adj.	original before	edited before	edited after	Iceland after
#	1551572	1400999	1400999	54370
mean	-0.34	-0.51	-0.97	0.46
RMS	18.01	10.23	7.78	4.20
min	-159.87	-155.63	-157.73	-60.52
max	221.25	115.55	106.00	40.56

RMS crossover difference before the adjustment reduces from 18.0 mgal to 10.2 mgal, solely through the editing of 3.6% bad data and the exclusion of the Red Sea sources (8.7%). A further reduction of the differences is resulting from the crossover adjustment, yielding a RMS value of 7.8 mgal. Furthermore, even better results can be obtained in areas with high quality shipborne gravity data, e.g., around Iceland a RMS value of only 4.2 mgal is found, cf. Table 4. However, it has to be noted that close to islands, where the altimeter results may have problems, larger RMS differences can be found, e.g., about 10 mgal around the Canary Islands. Similar results from the comparison of shipborne and altimetric gravity data were obtained by other researches, e.g., RMS differences of 8.5 mgal were found by Featherstone (2003), 7–11 mgal by Rapp (1998), 5.8 mgal by Andersen and Knudsen (1998), and 3–7 mgal by Sandwell and Smith (1997). At last, the differences between shipborne and altimetric gravity data are not displayed here due to space restrictions, but color plates of the differences between the unadjusted and crossover adjusted ship gravity data (edited data set) and KMS02 can be found in Denker and Roland (2003).

6 Conclusions

In this contribution, a consistent marine gravity data set for the European seas was compiled. This was achieved by merging data from four institutions, regeneration of the track structure, data editing, and crossover adjustment. Solely through the data editing, the RMS crossover difference reduced from 15.5 mgal to 8.4 mgal, while the crossover adjustment led to a final value of 4.7 mgal. Similarly, the differences between shipborne and altimetric data from KMS02 improved, yielding RMS values of 18.0 mgal for the original data, 10.2 mgal for the edited data before adjustment, and 7.8 mgal for the edited and crossover adjusted data. An even better agreement can be found in areas with high quality shipborne gravity data, e.g., around Iceland the RMS difference reduces to 4.2 mgal. The KMS02 altimetric data showed the best agreement with the compiled final ship data set with a RMS difference of 7.8 mgal, while the corresponding values are 8.2 mgal for both GSFC00 and Scripps and 9.2 mgal for CLS99.

Acknowledgements

We thank the German Research Foundation (DFG) for sponsoring this research under project DE 459/5-1/2, and we are grateful to the Bureau Gravimétrique International (BGI), the National Imagery and Mapping Agency (NIMA), and the National Geophysical Data Center (NGDC) for providing marine gravity data to us.

References

Andersen, O. B. and Knudsen, P. (1998). Global marine gravity field from the ERS-1 and Geosat geodetic mission altimetry. *Journal of Geophysical Research*, 103(C4):8129–8137.

Andersen, O. B., Knudsen, P., Kenyon, S., and Trimmer, R. (2003). KMS2002 Global Marine Gravity Field, Bathymetry and Mean Sea Surface. Poster, IUGG 2003, Sapporo, Japan, June 30-July 11, 2003.

Behrend, D. (1999). Untersuchungen zur Schwerefeldbestimmung in den europäischen Randmeeren. Nr. 229, Wiss. Arb. der Fachr. Verm.wesen der Univ. Hannover.

CLS (2003). *CLS – Mean Sea Surface Homepage*. Collecte Localisation Satellites. `http://www.cls.fr/html/oceano/projets/mss/welcome_en.html`.

Denker, H. and Roland, M. (2003). Compilation and Evaluation of a Consistent Marine Gravity Data Set Surrounding Europe. Poster, IUGG 2003, Sapporo, Japan, June 30-July 11, 2003. `http://www.ife.uni-hannover.de/download-allgemein/iugg03_marine.pdf`.

Featherstone, W. E. (2003). Comparison of Different Satellite Altimeter-Derived Gravity Anomaly Grids with Ship-Borne Gravity Data Around Australia. In Tziavos, I. N., editor, *Gravity and Geoid – 3rd Meeting of the International Gravity and Geoid Commission, Thessaloniki, Greece, August 26-30, 2002*, pages 326–331. Publishing Ziti.

GSFC (2003). *GSFC – Mean Sea Surface Height Homepage*. Goddard Space Flight Center. `http://magus.stx.com/mssh/mssh.html`.

Rapp, R. H. (1998). Comparison of Altimeter-Derived and Ship Gravity Anomalies in the Vicinity of the Gulf of California. *Marine Geodesy*, 21:245–259.

Sandwell, D. T. and Smith, W. H. F. (1997). Marine gravity anomaly from Geosat and ERS 1 satellite altimetry. *Journal of Geophysical Research*, 102(B5):10039–10054.

Strang van Hees, G. L. (1983). Gravity Survey of the North Sea. *Marine Geodesy*, 6(2):167–182.

Torge, W. (1989). *Gravimetry*. de Gruyter, Berlin, New York.

Wenzel, H. G. (1992). Sea gravity data adjustment with program SEAGRA. *Bureau Gravimétrique International, Bulletin d'Information*, 71:59–70.

Wessel, P. and Watts, A. B. (1988). On the Accuracy of Marine Gravity Measurements. *Journal of Geophysical Research*, 93(B1):393–413.

Comparison of Downward Continuation Methods of Airborne Gravimetry Data

F.Mueller, T. Mayer-Guerr
Institut of Theoretical Geodesy,
The University of Bonn, 53115 Bonn, Nussallee 17, Germany

Abstract. The determination of the geoid represents a central task in geodesy. In this matter airborne gravimetry provides a fast and reliable method for geoid determination especially for those regions, which are difficult to survey. In addition with respect to the spectral representation of the earth's gravity field in terms of spherical harmonics airborne gravimetry is capable to fill the gap with regard to the medium wavelength contribution. With current airborne gravity systems an accuracy of 2-3 mgal (rms), with a resolution of 6-15 km can be achieved. Data processing becomes more significant to meet the demands for a geoid with centimetre-accuracy. Besides the task of pre-processing which results in band-limited gravity disturbances at flight level, the choice of a proper downward continuation method gains in importance.

Besides the problem of regularization most calculations are based on block mean values, which has disadvantages in the approximation accuracy. This paper will compare two different approaches of downward continuation of simulated airborne gravimetry data. In the first approach gravity disturbances derived at flight altitude are directly transformed to the disturbing potential at geoid level. This method derives the disturbing potential in terms of block mean values, which are not continuously differentiable. With respect to that problem an alternative approach is proposed. This method will make use of space localizing spline functions to represent the disturbing potential at geoid level. It will be shown that this will increase the accuracy regarding the resulting geoid noise.

Keywords. Airborne Gravimetry, Downward Continuation

1 Introduction

Gravity observations on moving platforms provide continuous values along the survey profile. The aircraft observes the total acceleration $|\mathbf{g}|$, which is composed of the earth's gravity field $\mathbf{g}$ and other disturbing accelerations $\mathbf{a}$, caused by the aircraft's motion. This is illustrated in Figure 1.The major problem is to separate the desired gravity signal from the disturbing acceleration. With the help of kinematic differential GPS the position of the aircraft can be obtained and the disturbing accelerations can be removed at every point. To account for the still remaining noise, the data is low-pass filtered, resulting in band-limited gravity disturbances at flight altitude.

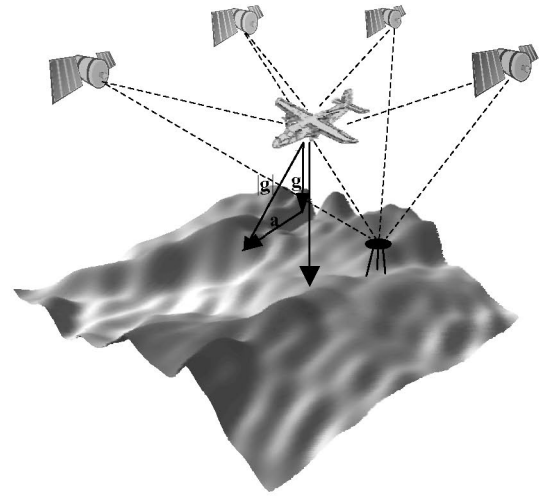


Fig.1 The basic principle of airborne gravimetry.

A main object in physical geodesy is to determine the geoid heights. For that reason the disturbing potential T(Q) at sea level has to be derived. The solution of that boundary problem involves integration, which is in most cases based on block mean values. A comparison between a method using block mean values and in contrast to that a method using space localizing spline functions for the case of downward continuation is treated in the following.

2 Theory

In this chapter the basic differences of block mean values and space localizing spline functions as base functions are treated in general.

2.1 Block mean values

Most integrals in physical geodesy are based on the Fredholm integral equation of first kind, which can be given in general form as,

$$V(P)=\frac{1}{4\pi}\iint_{\Phi}K(P,Q)f(Q)d\Phi(Q), \qquad (1)$$

where P relates to flight level and Q to geoid level respectively. The functionals f(t) can be described as,

$$f(Q)=\sum_{i=1}^{N}a_i\varphi_i(Q), \qquad (2)$$

where the $\varphi_I(Q)$ represent the kind of base functions. For block mean values these are given by,

$$\varphi_i(Q)=\begin{cases}1 : Q\in\Omega_i\\ 0 : Q\in\Phi\setminus\Omega_i\end{cases}, \qquad (3)$$

for the respective block Ω_i. Substituting equation (2) into (1) results in:

$$V(P)=\frac{1}{4\pi}\sum_{i=1}^{N}a_i\iint_{\Omega_i}K(P,Q)d\Omega_i(Q). \qquad (4)$$

In this case the remaining integral is approximated by the kernel's value at the centre of the block. It is obvious that this has a negative influence with regard to the approximation accuracy. Consequently, the functionals V(P) are calculated as follows:

$$V(P)=\frac{1}{4\pi}\sum_{i=1}^{N}a_i K(P,Q_i)\Delta\Omega_i \qquad (5)$$

2.1 Space localizing spline functions

The main characteristic of spline functions is their continuous form. A general form of the spline function can be given as:

$$\varphi_i(Q)=\sum_{n=0}^{\infty}k_nP_n(Q,S_i). \qquad (5)$$

The k_n define the shape of the spline function, in this particular case degree variances are used. The S_i are the points where the splines are set up and P_n represent the respective Legendre polynomials. The points S_i can be arranged in different ways, such as a triangular grid, which is realized in this case study. In analogy to the representation with block mean values V(P) reads:

$$V(P)=\frac{1}{4\pi}\sum_{i=1}^{N}a_i\iint_{\Phi}K(P,Q)\sum_{n=0}^{\infty}k_nP_n(Q,S_i)d\Phi(Q). \qquad (6)$$

In geodetic applications the kernel K(P,Q) can be developed in terms of Legendre polynomials P_n,

$$K(P,Q)=\sum_{n=0}^{\infty}(2n+1)d_n\left(\frac{R}{r}\right)^{n+1}P_n(P,Q), \qquad (7)$$

where the factor d_n defines the respective kernel. Inserting this kernel in equation (6) results in,

$$V(P)=\frac{1}{4\pi}\sum_{i=1}^{N}a_i\sum_{n=0}^{\infty}\sum_{\overline{n}=0}^{\infty}(2n+1)d_n\left(\frac{R}{r}\right)^{n+1}k_{\overline{n}}\times\iint_{\Phi}P_n(P,Q)P_{\overline{n}}(Q,S_i)d\Phi(Q), \qquad (8)$$

with

$$P_n(P,Q)=\frac{1}{2n+1}\sum_{m=-n}^{n}Y_{nm}(P)Y_{nm}(Q) \qquad (9)$$

and

$$\iint_{\Phi}Y_{\overline{nm}}(Q)Y_{nm}(Q)d\Phi(Q)=4\pi\delta_{n\overline{n}}\delta_{m\overline{m}}. \qquad (10)$$

Applying the addition theorem of spherical harmonics (9) and taking the orthogonality conditions (10) into account one finally gets:

$$V(P)=\sum_{i=1}^{N}a_i\sum_{n=0}^{\infty}d_n\left(\frac{R}{r}\right)^{n+1}k_n\frac{1}{2n+1}\sum_{m=-n}^{n}Y_{nm}(P)Y_{nm}(S_i)=\sum_{i=1}^{N}a_i\sum_{n=0}^{\infty}d_n\left(\frac{R}{r}\right)^{n+1}k_nP_n(P,S_i). \qquad (11)$$

As shown in equation (11) no approximation has

taken place with respect to the integration in contrast to the use of block mean values.
In the following this background is applied to the downward continuation of simulated airborne gravimetry data.

3 Numerical tests

3.1 Simulated dataset

The band-limited gravity distubances at flight level are derived from the ultra high degree geopotential model GPM3E97A (Wenzel,G. (1998)), which is tailored to Europe. They are computed as:

$$\delta g^b(P) = \frac{GM}{R^2}\sum_{n=l}^{l+b}(n+1)\left(\frac{R}{R+H}\right)^{n+2} T_n(P). \quad (12)$$

With respect to the spectral content of current airborne gravity data, the spectral band between degrees l=121 and l+b=1800 (maximum frequency) is used corresponding to a maximum resolution of 5 km half-wavelength. In addition a noise level of ± 1.5 mgal was added to the synthetic gravity data, see Figure 2. The data corresponds to a sampling rate of 2 arcmin resulting in 8281 gravity disturbances at flight level within an area of 48° - 51° in latitude and 10°-13° in longitude.

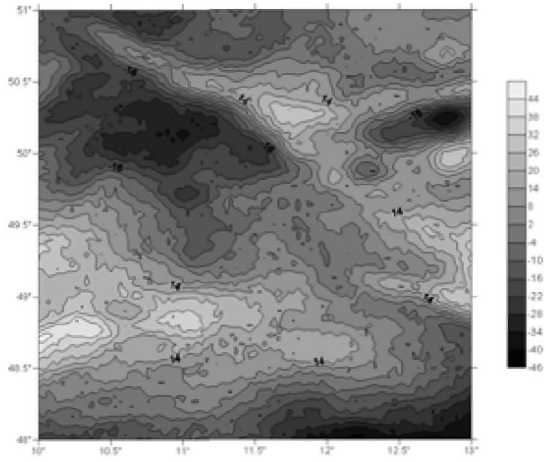


Fig. 2 Simulated gravity disturbances at 3km [mgal]

3.2 Approach 1

The main goal in geoid calculations is to determine the geoid heights. For that reason the disturbing potential T(Q) at sea level has to be derived. In spherical approximation, the problem is formulated as follows (Novak, Heck (2002)) :

$$\begin{aligned} &\Delta T^b(P) = 0 && for \quad r > R \quad (13)\\ &\delta g^b(P) = -\left.\frac{\partial T^b(r,\Omega)}{\partial r}\right|_r && for \quad r = R+H\\ &\lim_{r\to\infty} T^b(P) = 0 \end{aligned}$$

In the case of airborne gravimetry this is an improperly posed problem. The following approach takes advantage of the typical character of airborne collected gravity data, which are:

- band-limited and
- measured at almost constant height.

The improperly posed problem is transformed to a properly posed problem by regularization. This results in a direct solution, where the gravity disturbances at flight level are directly transformed to the disturbing potential at sea level. The disturbing potential at geoid level is given by:

$$T^b(Q) = \frac{R+H}{4\pi}\iint_{\Phi} K^b(P,Q)\delta g^b(P)d\Phi. \quad (14)$$

The kernel K is represented by :

$$K^b(P,Q) = \sum_{n=l}^{l+b}\frac{2n+1}{n+1}\left(\frac{R+H}{R}\right)^{n+1} P_n(P,Q). \quad (15)$$

The integration is carried out over a spherical cap of radius $\Psi_0 = 1°$. In this context, the integration is divided into a near zone and a far zone contribution. All points within the spherical cap belong to the near zone contribution. The far zone contribution is modelled with the help of a GGM (Global Geopotential Model) such as the EGM96 up to degree L=360. Consequently, the discretized integral reads,

$$\begin{aligned} &T^b(Q_j) - \frac{GM}{2(R+H)}\sum_{n=l}^{L\le l+b}(n+1)\left(\frac{R}{R+H}\right)^{n+1} V_n(H,\Psi_0)T_n(Q_j) \quad (16)\\ &= \frac{R+H}{4\pi}\sum_{i=1}^{N}\delta g^b(P_i)K^b(P_j,Q_i)\Delta\Omega_i \ , \end{aligned}$$

where V_n are the truncation coefficients,

$$V_n(H,\Psi_0) = \sum_{m=l}^{l+b}\frac{2m+1}{m+1}\left(\frac{R+H}{R}\right)^{m+1} R_{n,m}(\Psi_0) \quad (17)$$

and the $R_{n,m}$ represent the Paul-coefficients. They can be computed iteratively, see Paul (1973).
In matrix form equation (16) reads,

$$\mathbf{T}^b(Q_j) = \frac{R+H}{4\pi}\mathbf{A}\,\delta\mathbf{g}^b(P_i), \tag{18}$$

and the matrix elements are given as:

$$A_{ij} = \begin{cases} K^b(P_i,Q_j)\Delta\Omega_i & for \quad \angle(P_i,Q_j) \le \Psi_0 \\ 0 & for \quad \angle(P_i,Q_j) > \Psi_0 \end{cases} \tag{19}$$

and

$$A_{jj} = 2\pi \int_{\Psi=0}^{\Psi_0} K(\Psi)\sin\Psi\, d\Psi - \sum_{i=1,i\neq j}^{N} K^b(P_i,Q_j)\Delta\Omega_i. \tag{20}$$

For further details see also Novak et al. (2001) or Kern (2003).
The calculations are bounded by 49°-50° in latitude and 11°-12° in longitude in order to avoid any edge effects. With a sampling rate of 2 arcmin this will result in 600 computation points. Figure 3 illustrates the differences between the true disturbing potential and the disturbing potential derived by integration in terms of the geoid noise calculated with Bruns formula. The displayed results are related to a flight altitude of 3 km.

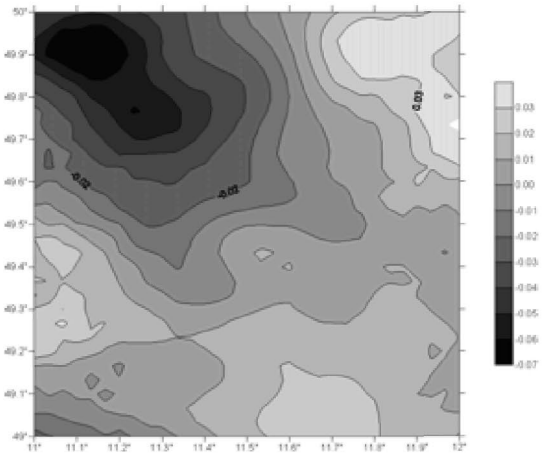


Fig. 3 Geoid noise using approach 1 [m]

The results for different flight levels are shown in table 1.

Table 1 Geoid noise for different heights (approach 1) [cm]

Height (km)	Minimum	Maximum	Mean	RMS
3	-6.4	+3.5	-2.0	±2.2
4	-6.8	+3.7	-2.1	±2.2
6	-7.7	+4.2	-2.3	±2.4

3.2 Approach 2

The functionals in method 1 are represented in terms of block mean values. With respect to equation (11) the following approximation approach will be used :

$$T(Q) = \sum_{i=1}^{N} a_i \sum_{n=0}^{\infty} k_n\, P_n(Q,S_i). \tag{21}$$

Thus, the gravity disturbances at flight level can be expressed by the upward continuation using the Neumann function:

$$\delta g(P) = \frac{R}{4\pi}\iint_{\Phi} K(P,Q)_{Neumann}\, T(Q)\, d\Phi. \tag{22}$$

Substituting the approximation approach (21) in (22) results in:

$$\delta g(P) = \frac{1}{4\pi}\iint_{\Phi}\left[\sum_{n=0}^{\infty}(2n+1)(n+1)\left(\frac{R}{R+H}\right)^{n+2} P_n(P,Q)\right] \times \underbrace{\left[\sum_{i=1}^{N} a_i \sum_{n=0}^{\infty} k_n P_n(Q,S_i)\right]}_{T(R,\Omega_j)} d\Phi. \tag{23}$$

After applying the addition theorem for spherical harmonics this equation reads:

$$\delta g^b(P) = \sum_{i=1}^{N} a_i \sum_{n=l}^{l+b}(n+1)\left(\frac{R}{R+H}\right)^{n+2} k_n P_n(P,S_i). \tag{24}$$

This represents the observation equation, in which the unknown spline parameters a_i are estimated. Figure 4 shows the process of computing the regularization parameter by using variance components estimation. To prevent any edge effects the area was bounded in analogy to method 1, resulting in n_{obs}= 8281 observations and u=1133 unknowns. The contribution of EGM96 (Lemoine et al., 1998) of degree n=121-360 was firstly removed and afterwards added to the results, which

corresponds to the classical remove and restore technique.

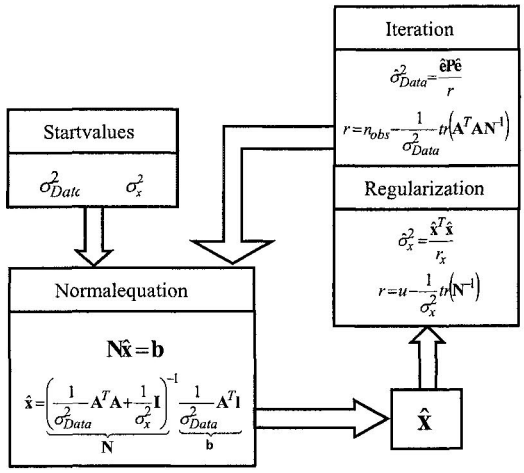


Fig. 4 Variance components estimation

The geoid noise in analogy to approach one is displayed in Figure 5. Results for different heights are listed in table 2.

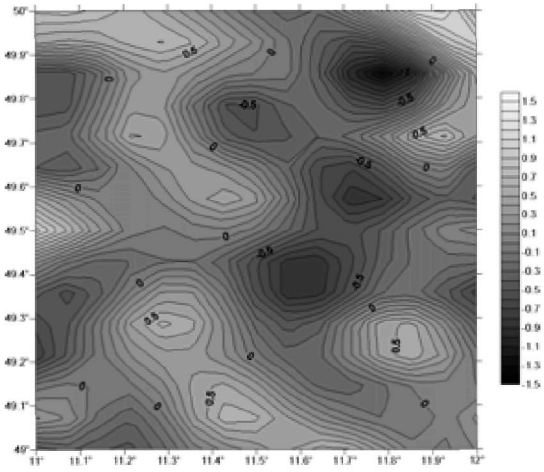


Fig. 5 Geoid noise using approach 2 [cm]

Table 2 Geoid noise for different heights (approach 2) [cm]

Height (km)	Minimum	Maximum	Mean	RMS
3	-1.4	+1.3	+0.4	±0.4
4	-1.5	+1.6	+0.4	±0.5
6	-2.0	+2.2	+0.5	±0.7

3 Conclusions

It has been demonstrated that the results by using space localizing spline functions perform better in terms of accuracy (see table 1, table 2). At a flight altitude of 3km the rms using approach 1 is ±2.2cm in contrast to approach 2 with ±0.4cm. Further improvements may be achieved by using different kinds of base functions such as wavelets. Modifications of the spline function, particularly with regard to the coefficients k_n which define the shape of the spline function, could also increase the accuracy.

References

Kern M (2003) An analysis of the combination and downward continuation of satellite, airborne and terrestrial gravity data. UCGE Tech. Rep. 20172 The University of Calgary.

Novak P, Heck B (2002) Downward continuation and geoid determination based on band-limited airborne gravity data. Journal of Geodesy 76:269-278

Lemoine FG and 14 authors (1998) The Development of the Joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM96, NASA/TP-1998-206861, NASA-GSFC, Greenbelt MD

Novak P, Kern M, Schwarz KP, Heck B, (2001) On the determination of a band-limited gravimetric geoid from airborne gravimetry. UCGE Tech Rep 30013, The University of Calgary

Paul M (1973) A method of evaluation the truncation error coefficients for geoidal heights Bull. Géd 110: 413-425

Wenzel G (1998) Ultra-high degree geopotential models GPM98 A, B and C to degree 1800. Proc of the joint meeting Int. Gravity Commission and Int Geoid Commission, Trieste, 7-12 September (www.gik.uni-karlsruhe.de)

Optimisation of Density Parameter in Gravity Prediction from Bathymetry

D.M. Sproule, A.H.W. Kearsley
School of Surveying and Spatial Information Systems,
University of New South Wales, Sydney, 2052, Australia
d.sproule@student.unsw.edu.au

Abstract. This paper describes experiments conducted to examine the combination of gravity values predicted from bathymetric data with observed gravity data to increase the resolution of offshore gravity data. The density parameter in the prediction is optimised empirically.

The gravity effect resulting from bathymetric features can provide short wavelength features of the gravity field, but at longer wavelengths geophysical features below the seafloor begin to influence the gravity signal. A model of the isostatic response might help to determine these longer wavelength features, but these models are inconclusive so in this case the calculations are restricted to the bathymetry.

By using a combination of observed gravity and gravity predicted from bathymetry, the long wavelength features present in the direct observations can be preserved while using the bathymetric data to fill-in the short wavelength features.

The computation scheme is described, and some numerical results are presented. The effect of increasing the distance between the observed gravity data is examined.

For the calculation of the gravity signal arising from the bathymetry, a single density value is chosen for the entire area. Due to density variations this density may not be optimal. Different density values are tested to examine the effect this choice has on the resulting predicted gravity.

Keywords. Bathymetry, gravity prediction

1 Introduction

In offshore areas where the availability of gravity data is limited, it may be possible to utilise bathymetric information to assist the densification of the gravity data.

Bathymetric information on its own cannot provide knowledge of the entire gravity signal; however it may be possible to use bathymetric prediction of gravity in conjunction with observed gravity.

This has applications for the densification of gravity data, obtained from airborne surveys for instance.

2 Mathematical formulation of the problem

2.1 Gravity effect of a layer

In the spatial domain, the gravitational effect at a point x,y,z of a topographic layer bounded by z_1 and z_2 is given by a triple integral:

$$g_z = -\gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{z_1}^{z_2} \rho(x',y')\frac{(z-z')}{r^3}dz'dx'dy' \qquad (1)$$

,where γ is the gravitational constant, and

$$r^2 = (x-x')^2 + (y-y')^2 + (z-z')^2. \qquad (2)$$

Parker (1972) describes a computationally efficient method for calculating this, which uses Fourier transforms to efficiently deal with the integrals involved. The resulting equation, as summarised by Blakely (1995), is:

$$\Im[g_z] = 2\pi\gamma e^{|k|z_0}\sum_{n=1}^{\infty}\frac{(-|k|)^{n-1}}{n!}\Im\left[\rho\left(z_1^n - z_2^n\right)\right] \qquad (3)$$

where g_z is the gravitational effect, at a height z_0, of a layer bounded by the upper and lower surfaces $z_1(x,y)$ and $z_2(x,y)$ respectively, and having the density function $\rho(x,y)$. The notation $\Im[\ ]$ indicates Fourier transforms. γ is the gravitational constant, 6.67×10^{-11} m^3 kg^{-1} s^{-2}.

In this study, z_1 corresponds to the seafloor and is given by the bathymetric information. The surface z_2 is not directly known and depends on the isostatic response. Existing isostatic models are only partially successful and in the absence of any direct knowledge about the isostatic response, it is chosen to consider only that surface which has actual observed data, that is the bathymetric data.

Thus, the formula reduces to the following (see Blakely 1995):

$$\Im[g_z] = 2\rho\pi\gamma e^{|k|z_0}\sum_{n=1}^{\infty}\frac{(-|k|)^{n-1}}{n!}\Im\left[z_1^n\right] \qquad (4)$$

The density function in this case is also assumed constant in the absence of any additional information. The mean crustal density value of 2.67 g cm^{-3} and 1.027 g cm^{-3} for seawater are used, giving a density change of 1.643 g cm^{-3}. This value is used in the calculations initially, and subsequently different values are trialed to examine the effect.

Obviously, this cannot provide total information about the gravity signal. Fortunately, at short wavelengths, the isostatic compensation signal at the surface is negligible because it is strongly attenuated (see McNutt 1979).

This means that by utilising the bathymetric effect in the short wavelength domain only, the unknown compensation effect can by largely ignored. To this end, the bathymetric information will be limited to providing 'fill-in' data to determine gravity data between existing observed gravity.

The long wavelength features in the gravity signal arising from the unmodelled isostatic response should be present in the observed gravity data as long as the spacing of these observations is not too large. The aim of the computations in this paper is to examine how dense the observed gravity data should be to avoid large errors in the predicted gravity signal.

Ultimately, the processing scheme is to use observed gravity at a certain spatial resolution; and then incorporate the bathymetric information to densify the gravity information.

2.2 Processing scheme

Firstly, gridded bathymetric data is used to produce a grid of the gravity signal arising as the gravimetric

response of that bathymetry. This quantity $\Delta g^{BATHYMETRIC}$ is the inverse fourier transform of equation (2).

The bathymetric effect is then subtracted from the observed gravity anomalies. This residual can be called the 'non-bathymetric' effect.

$$\Delta g' = \Delta g^{OBSERVED} - \Delta g^{BATHYMETRIC} \quad (5)$$

Note that this computation is performed at the points of the observed gravity data, with a thinned set of $\Delta g^{BATHYMETRIC}$ produced by linear interpolation.

The residual is then interpolated to a grid at the final required density ($\Delta g'^{GRIDDED}$).

The bathymetric effect is then added back to the interpolated data.

$$\Delta g^{PREDICTED} = \Delta g'^{GRIDDED} + \Delta g^{BATHYMETRIC} \quad (6)$$

This predicted gravity field combines the information from the observed gravity data and from the effect calculated from the bathymetric data.

The result is to use the bathymetric effect to interpolate between the observed gravity data; the observed gravity data provides the long-wavelength component, while the bathymetric effect provides the short wavelength component interpolated between observed gravity points.

3 Data sources

The bathymetric data used is the International Bathymetric Chart of the Arctic Ocean (IBCAO) (see Jakobsson et. al. 2000).

The area chosen for the tests was bounded by 71° $\leq \phi \leq$ 80°, and 345° $\leq \lambda \leq$ 5°.

For this study, the dataset was interpolated to a sparser resolution of 0.05 degrees latitude (approx. 5.5 km) by 0.2 degrees longitude (averaging approx. 5 km). The bathymetry in this area has a mean depth of 2109 m and the magnitude of the bathymetry has an rms of 1185 m.

The gravity data (free air anomalies) was obtained from the KMS database. The data was extracted at a resolution to match the resolution of the bathymetric chart data. The data has the following characteristics : a mean of 2 mgal and an rms of 20 mgal, with a minimum value of -120 mgal, and maximum value of 112 mgal.

4 Data processing

The processing was performed using GRAVSOFT software developed at KMS by R. Forsberg.

Equation 2 is an infinite expansion which has to be truncated to allow efficient computations. The computation of the bathymetric effect used 5 terms of the expansion, as the calculation was seen to converge at the sub-milligal level with this number of terms.

The gravity data was extracted from the KMS database at a resolution of 0.2 degrees longitude and 0.1 degrees latitude. This gravity data at the highest density formed the test dataset against which the prediction technique was compared (Δg^{TEST}).

The 'observed' gravity data used for the prediction was extracted from the same source but interpolated at lower resolution.

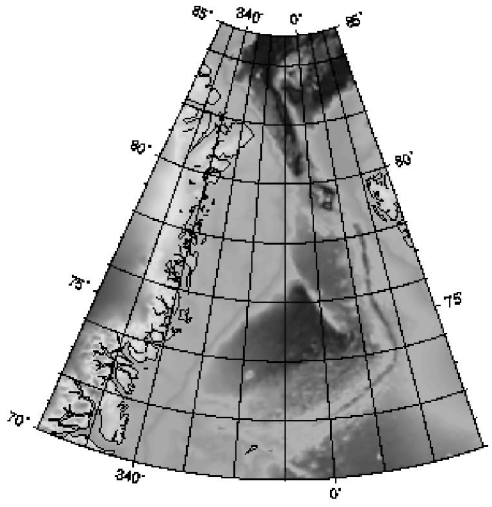


Fig. 1. Location of study area showing bathymetric data

Logically the predicted gravity would be expected to be less accurate if the resolution of the observed gravity data is less. Hence, gravity data sets were created at various grid spacings from the original data. These were used to simulate gravity data observed at varying spatial resolutions ($\Delta g^{OBSERVED}$).

The spacing of these 'observed' data sets was varied by varying the longitude spacing from 0.2 degrees up to 1.7 degree longitudes with 0.1 degree steps, with the latitude spacing of 0.05 degrees remaining constant.

The algorithm as described earlier, was employed to produce a 'predicted' gravity field from each of the 'observed' gravity sets.

To test the quality of the prediction it was compared against the test gravity set by computing the error:

$$\varepsilon = \Delta g^{TEST} - \Delta g^{PREDICTED} \quad (7)$$

where ε was computed for each of the 'observed' gravity data sets. This allows an assessment of how the error increases with the decreasing resolution of the 'observed' gravity data.

Using the same gravity data sets, a linear interpolation was also performed without using the bathymetric information. i.e. $\Delta g^{INTERPOLATED}$ is $\Delta g^{OBSERVED}$, interpolated to the grid spacing of the test gravity data. These results were also compared to the test gravity data.

$$\varepsilon = \Delta g^{TEST} - \Delta g^{INTERPOLATED} \quad (8)$$

5 Varying the density parameter

Initially the mean values of 2.67 g cm^{-3} for crust and 1.027 g cm^{-3} for salt water were used. The reults are shown in Table 1, which also shows the result from simple interpolation.

Without specific knowledge of the crustal density in this region, the chosen value may not be appropriate. The value for the density parameter can be determined by the inverse method using observed gravity data and the bathymetric information, assuming that both these data sets are available with good spatial resolution. Since the aim of the prediction technique is to increase the spatial resolution of observed gravity data where it is not present at the desired resolution, it would become a circular argument to attempt to determine the crustal density parameter first from existing gravity and bathyemetry data and then use this parameter to predict the gravity signal.

As an alternative, different trial values were chosen for crustal density to examine how sensitive the prediction was to this parameter. The range chosen for the crustal density was between 2.30 and

2.90 g cm^{-3}, in 0.1 g cm^{-3} steps. For each of the values in this range the procedure described in Section 4 was followed.

The results are tabulated in Table 2.

Table 1. RMS of ε

Longitude spacing (degrees)	RMS of ε in $\Delta g^{PREDICTED}$ (mgal)	RMS of ε in $\Delta g^{INTERPOLATED}$ (mgal)
0.2	0.00	0.00
0.3	1.21	1.26
0.4	1.76	1.87
0.5	2.51	2.66
0.6	3.27	3.52
0.7	4.03	4.31
0.8	4.82	5.22
0.9	5.51	5.97
1.0	5.88	6.43
1.1	6.70	7.31
1.2	7.56	8.32
1.3	7.90	8.76
1.4	8.83	9.66
1.5	9.12	10.12
1.6	9.80	10.79
1.7	10.17	11.26

Table 2. RMS of ε in $\Delta g^{PREDICTED}$ (mgal) for different assumed densities

	Assumed crustal density (g cm^{-3})						
Longitude spacing (degrees)	2.30	2.40	2.50	2.60	2.70	2.80	2.90
0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.3	1.21	1.21	1.21	1.21	1.21	1.22	1.22
0.4	1.77	1.77	1.77	1.76	1.76	1.76	1.76
0.5	2.52	2.51	2.51	2.51	2.51	2.51	2.51
0.6	3.29	3.29	3.28	3.27	3.27	3.27	3.27
0.7	4.05	4.04	4.03	4.03	4.03	4.02	4.03
0.8	4.86	4.84	4.83	4.83	4.82	4.82	4.82
0.9	5.55	5.53	5.52	5.52	5.51	5.51	5.51
1.0	5.92	5.91	5.90	5.89	5.88	5.88	5.88
1.1	6.74	6.72	6.71	6.70	6.70	6.70	6.71
1.2	7.62	7.60	7.58	7.57	7.56	7.56	7.56
1.3	7.96	7.94	7.92	7.90	7.89	7.89	7.90
1.4	8.86	8.85	8.83	8.83	8.83	8.85	8.86
1.5	9.16	9.14	9.13	9.12	9.12	9.13	9.15
1.6	9.82	9.80	9.79	9.79	9.80	9.82	9.84
1.7	10.20	10.18	10.17	10.17	10.18	10.20	10.23

6 Results

Table 1 shows the rms of the error in the prediction, ε, measured over all points in the test area, using the standard density value. The prediction technique using the bathymetric information shows a trend of being constantly better than the simple linear interpolation.

The simple interpolation of results was outperformed by about ten percent.

The results in Table 2 show the rms of error when using different assumed density values.

The table shows that the density assumed can affect the results. The initially chosen value of 2.67 produces the best results. Small variations in the density value did not significantly affect the result, so that choosing any value between 2.50 and 2.67 for example had very little effect on the results. The results degraded as the density values departed more from the original value. The differences are most obvious in the last row of the table, which corresponds to the most widely spaced 'observed' data points. It can be noted that although there is a trend for the error to increase as the chosen density departs further from 2.67, the results are identical at the level of 0.1 mgal. This would seem to indicate that the computation is not particularly sensitive to the choice of crustal density.

These results show that the standard value of 2.67 is appropriate in this region of study, but this may not be so for other regions. A further shortcoming is that the 'best' value for crustal density is determined here by a comparison with a spatially dense observed gravity field which of course is not available in a region where the intention is to densify a sparseley observed gravity field.

7 Conclusion and future possibilities

Although bathymetric data on its own cannot provide full knowledge of the gravity field, it can be used alongside existing gravity measurements to improve the knowledge of the gravity field. Measurable improvement can be achieved by incorporating bathymetric data into observed gravity data, when compared with simply interpolating the existing gravity data.

This technique could prove useful for improving geoid models from areas with less dense observed gravity data. Further computations could show the effect of this error on using this data in geoid computations.

The standard mean value for crustal density did produce in this case the best results. No improvement in the result was achieved by varying the density in the computation, and in fact the computation seemed fairly insensitive to the choice of density value.

This outcome will not necessarily be valid for all regions, and the method of choosing an appropriate value for crustal density for a particular region by comparing the results to the known gravity field will not be available where gravity observations are not present at high resolution. Since this prediction technique might be useful in just such regions of sparseley observed gravity, in can only be said that the standard value for crustal density may be the most appropriate value to use where no additional information is available.

References

Blakely RJ (1995) *Potential theory in gravity and magnetic applications*, Cambridge University, p295.

Jakobsson M, Cherkis NZ, Woodward J, Macnab R, and Coakley B (2000) New grid of Arctic bathymetry aids scientists and mapmakers; *Eos*, Transactions, American Geophysical Union, v. 81, no.9, p.89,93,96.

McNutt M (1979) Compensation of Oceanic Topography: An Application of the Response Function Technique to the 'Surveyor' Area, *J. Geophys. R., 84*(B13,10 December), 7589-7598,.

Parker RL (1972) The rapid calculation of potential anomalies, *Geophysical Journal of the Royal Astronomical Society* 31, pp. 447-455gv

Calibration and Validation of GOCE Gravity Gradients

J. Bouman*[1], R. Koop[1], R. Haagmans[2], J. Müller[3], N. Sneeuw[4], C.C. Tscherning[5], and P. Visser[6]

[1]SRON National Institute for Space Research, Utrecht, The Netherlands
[2]European Space Agency, ESTEC, Noordwijk, The Netherlands
[3]Institut für Erdmessung, University of Hannover, Hannover, Germany
[4]Department of Geomatics Engineering, University of Calgary, Calgary, Canada
[5]Department of Geophysics, University of Copenhagen, Copenhagen, Denmark
[6]DEOS, Delft University of Technology, Delft, The Netherlands

*j.bouman@sron.nl

Abstract. GOCE will be the first satellite ever to measure the second derivatives of the Earth's gravitational potential in space. With these measurements it is possible to derive a high accuracy and resolution gravitational field if systematic errors have been removed to the extent possible from the data and the accuracy of the gravity gradients has been assessed. It is therefore necessary to understand the instrument characteristics and to setup a valid calibration model. The calibration parameters of this model could be determined by using GOCE data themselves or by using independent gravity field information. Also the accuracy or error assessment relies on either GOCE or independent data. We will demonstrate how state-of-the-art global gravity field models, terrestrial gravity data and observations at satellite track crossovers can be used for calibration/validation. In addition we will show how high quality terrestrial data could play a role in error assessment.

1 Introduction

The main goal of the GOCE mission (expected to be launched early 2006) is to provide unique models of the Earth's gravity field and of its equipotential surface, as represented by the geoid, on a global scale with an accuracy of 1 cm at 100 km resolution (ESA, 1999). To this end, GOCE will be equipped with a GPS receiver for high-low satellite-to-satellite tracking (SST-hl), and with a gradiometer for observation of the gravity gradients (SGG). The gradiometer consists of six 3-axes accelerometers mounted in pairs along three orthogonal arms. The sum of the accelerations of two accelerometers along one arm gives the so-called common mode (CM), which is the signal that is common to both accelerometers, atmospheric drag for example, while the difference between the accelerations of two accelerometers along one arm gives the differential mode (DM), which includes, for example, gravity gradients. The accelerometers are designed such as to give the highest achievable precision in the measurement bandwidth (MBW) between 5 and 100 mHz, which corresponds to a resolution along track of 40-8000 km. For the diagonal gravity gradients in a Local Orbital Reference Frame (LORF, x-axis in the velocity direction, the z-axis approximately radially outward and the y-axis complements the right-handed frame) this precision will not exceed 4 mE/$\sqrt{\text{Hz}}$ (1 E = 10^{-9} s^{-2}) in the MBW (ESA, 1999).

The observations will be contaminated with stochastic and systematic errors. For the GOCE gradiometer, systematic errors typically are due to instrument imperfections like misalignments of the accelerometers, scale factor mismatches etc. The CM and DM couplings, which are the result of such instrument imperfections, can be determined with a relative accuracy of $10^{-2} - 10^{-4}$ prior to the mission by the so-called pre-flight calibration on ground using a test bench. In orbit, a so-called internal (in-flight) calibration procedure will be used (ESA, 1999), by which the CM and DM couplings can be determined to an accuracy level at which their effect on the gradients in the MBW stays below the required 4 mE/$\sqrt{\text{Hz}}$.

The gravity gradients are derived from the internally calibrated DM accelerations. The internal calibration, however, is not sensitive to all instrument imperfections. The readout bias, for example, and the accelerometer mis-positioning cannot be accounted for. Therefore, in order to possibly correct for remaining errors after internal calibration, a third calibration step is required which is called external calibration (or absolute calibration). It is performed during or after the mission and typically makes use of external gravity data, e.g. terrestrial gravity data or global gravity field models (Arabelos and Tscherning, 1998; Koop et al., 2002).

Along with the external calibration of the observations, the error of the observations needs to be assessed. For this purpose we could again use terrestrial gravity data. Alternatively, we could use the GOCE data themselves and perform an internal assessment, see e.g. (Albertella et al., 2000; Koop et al., 2002). Bouman and Koop (2003a) used along track interpolation of gravity gradients, which seems to be an adequate tool for error assessment in the MBW.

Validation can be defined as the process of assessing, by independent means, the quality of the data products derived from the system outputs. In contrast to calibration, there

is no active correction applied to the data. With validation one could test the success of the calibration. One could validate the gravity gradients themselves (validation on Level 1), or one could validate derived products such as harmonic coefficients, gravity anomalies or geoid heights (validation on Level 2).

In Section 2 we give some typical examples of external calibration, while validation examples are given in Section 3.

2 External calibration

The calibration model we use should be related to the gravity gradient errors. Both accelerometer scale factor errors and biases result in equivalent gravity gradient errors, whereas both accelerometer non-orthogonality and misalignment erroneously project part of the signal (gravity gradients plus satellite rotation) onto the measured accelerations. Because the signal has most of its power at 1 and 2 cycles per revolution (cpr), the DM acceleration errors and the gravity gradient errors will show relatively large 1, 2, ... cpr errors. The time series of GOCE gravity gradients $\underline{y}_g$ may therefore be written as

$$\underline{y}_g(t) = \lambda[y_c(t) + \Delta y + y' \cdot t + \sum_k (a_k \cos k\omega(t) + b_k \sin k\omega(t))] + \underline{n}(t) \quad (1)$$

with scale factor λ, bias Δy, trend y' and Fourier coefficients a_k, b_k at $k = 1, 2, \ldots$ cpr as calibration parameters. The trend is included to model, for example, slow bias variations in time. In addition, t is time, ω frequency and $\underline{n}$ noise. The true gravity gradients y_c are of course unknown, but we will use calibration gradients instead, that is, approximate gradients $\underline{y}_c$ are computed using, for example, an existing global gravity field model.

Because the errors at long wavelengths of gravity gradients computed with a global gravity field model are expected to be smaller than the errors in the GOCE gravity gradients, it is fair to expect that external calibration using global models could remove systematic long wavelength errors from the gravity gradient time series. Hence, the calibration parameters can be estimated using measured gradients $\underline{y}_g$ and calibration gradients $\underline{y}_c$. Of course, the GOCE gradients are expected to be more accurate at short wavelenghts.

2.1 Global gravity field models

The results presented in this Section are based on the IAG SC7 test data set[1], see also (Bouman and Koop, 2003b). A 30 day GOCE gravity gradient time series was computed using EGM96 complete up to degree and order 300, and simulated errors were added. The interval between two consecutive measurements is 5 s. Calibration gradients were computed using OSU91A also complete up to degree and order 300 and a 'GRACE' gravity field model, that is, EGM96 but complete up to degree and order 120.

[1] See www.geod.uni-bonn.de/SC7/index.html

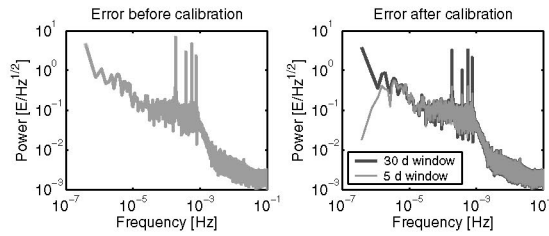


Fig. 1. PSD of V_{xx} gravity gradient errors before (external) calibration (left panel). Errors after calibration with OSU91A for a calibration window of 30 days and windows of 5 days (right panel).

In Figure 1 (left) the PSD of the V_{xx} errors is shown. Clearly, the errors at 1, 2, 3 and 4 cpr are large (peaks between 10^{-4} and 10^{-3} Hz), and in addition the errors on the low frequencies between 0 and 1 cpr are large. The former are caused by the coupling between the error and the signal, which is large at 1 and 2 cpr due to the Earth's flattening and the orbital eccentricity, while the latter are due to the accelerometer instability for lower frequencies. The V_{zz} errors show similar behaviour, while the V_{yy} errors at 3 and 4 cpr are relatively small (not shown).

The estimated calibration parameters are a scale factor, a bias and a trend for the total data set (30 days) for each of the three diagonal gradients, see model (1). In addition, Fourier coefficients at 1, 2, 3 and 4 cpr for V_{xx}, V_{zz}, and at 1 and 2 cpr for V_{yy} were estimated. The error reduction of the calibrated gradients with respect to the uncalibrated measurements is shown in Table 1. Calibration with the 'GRACE' model is only marginally better then calibration with OSU91A.

Although the gravity gradient errors have been reduced, we are now confronted with at least two limitations of our calibration model. The first is that errors at long wavelengths between 0 and 1 cpr other than a scale factor error and a trend are uncalibrated because in (1) only a scale factor and a trend have effect at these low frequencies. Secondly, the gravity gradient errors are correlated with the orbital height variation, which changes slowly but significantly with time. Hence, the signal and the errors at 1, 2, ... cpr vary in time as well, and the estimated Fourier coefficients are not representative for the whole 30 day period. The error reduction at 1,2, ... cpr is therefore modest (see Figure 1, right). One way to deal with these problems is to divide the calibration period

Table 1. Standard deviation of the errors after calibration with OSU91A. The numbers give the fraction with respect to the standard deviation of the errors before calibration. The numbers for calibration with 'GRACE' gradients are given between brackets. Besides a scale factor, a bias and trend, V_{xx} and V_{zz} are calibrated with 4 cpr coefficients, while V_{yy} is calibrated with 2 cpr coefficients.

window	V_{xx}	V_{yy}	V_{zz}
1×30 days	0.64 (0.64)	0.84 (0.80)	0.54 (0.53)
6×5 days	0.29 (0.29)	0.54 (0.42)	0.32 (0.27)

into shorter calibration windows and to estimate new calibration parameters for each observation window. If the window is small enough, then the estimated Fourier coefficients are representative for this window. In addition, this windowing will remove part of the long wavelength errors below 1 cpr: if calibration gradients are very accurate at the long wavelengths, then the window size determines the longest possible wavelength of the errors in the calibrated gradients.

The total observation period of 30 days has been divided into six windows of five days. For each of these windows we determined the full set of calibration parameters, that is, scale factor, bias, trend and up to 2 cpr Fourier coefficients for V_{yy} and up to 4 cpr Fourier coefficients for the other two gradients. Results for V_{xx} are shown in Figure 1, while Table 1 summarizes the reduction of the standard deviation of the errors with respect to the original uncalibrated gradients. As expected the long wavelength errors are reduced using shorter calibration periods. A period of 5 days, for example, corresponds to 2.3×10^{-6} Hz, and it is clear from Figure 1 (right panel) that the errors below this frequency have indeed become smaller. Furthermore, the errors at 1-4 cpr have been largely reduced. Calibration with the 'GRACE' model is better then calibration with OSU91A, except for V_{xx}. The latter is probably due to our accurate knowledge of the zonal harmonics to which especially V_{xx} is sensitive.

The above examples indicate that the long wavelength error of existing high resolution global gravity field models such as OSU91A or EGM96 is small enough to be used in the external calibration of GOCE gravity gradients. Furthermore, the calibration results may improve when more accurate CHAMP or GRACE based high resolution global gravity field models become available. The CHAMP-only and GRACE-only models that are available today are not suited for external calibration of the GOCE gradients. These models have full power approximately up to degree and order 65 and 90 respectively, whereas a proper calibration model needs to be complete, with full power, up to degree and order 120 at least. Otherwise the omission error is too large and there will be no improvement in the external calibration compared to, for example, EGM96.

2.2 Terrestrial data

Gravity gradients in the LORF may be predicted from terrestrial gravity data using, for example, least-squares collocation (LSC), (Tscherning, 1993). We use the GRAVSOFT package in our computations (Tscherning, 1974). The idea is to use gravity anomalies Δg to predict gradient anomalies T_{xx}, T_{yy} and T_{zz} in the LORF in a region with extension $(-125° \leq \lambda \leq -115°, 52° \leq \phi \leq 67°)$.

The SC7 data set, however, is not particularly suited for regional applications. We would like to have more high frequency signal and we therefore generated gravity gradients using EGM96 up to degree and order 360. In addition, the sampling was increased from 5 s to 1 s. The selected area contained 55 GOCE tracks, with approximately 1500 points. Simulated errors were added to the GOCE gradients, and the error RMS is 3-4 mE for the three different gradients.

Table 2. Statistics of the errors in the LSC predicted gravity gradients (1528 data points): predicted minus true. Gravity anomalies with 1 mGal error were used in the prediction (8601 data points). Between brackets are the LSC predicted standard deviations. Units are [mE].

gradient	mean	σ	min	max
T_{xx}	+0.1	0.6 (0.6)	-1.6	1.7
T_{yy}	+0.2	0.5 (0.7)	-1.3	1.2
T_{zz}	-0.3	0.9 (1.0)	-2.0	2.4

Secondly, 'true' gravity anomalies were computed using EGM96, $n = 51 - 360$. Because regional data cannot represent long wavelengths, we subtracted $n \leq 50$, which is on the safe side given the size of the region (corresponds to $n = 18$). The gradient prediction becomes more reliable by using data in a larger region, so we selected gravity anomaly data points on a grid with $-135° \leq \lambda \leq -105°$, $42° \leq \phi \leq 77°$, and $\Delta\lambda = 1/2°, \Delta\phi = 1/4°$ (8601 points in total). Errors were added to these data which were uncorrelated with zero mean and a standard deviation of 1 mGal. An empirical covariance function of the noisy gravity data was determined and an analytical model was fitted to this empirical function (Knudsen, 1987). The variance of the gravity data is 207 mGal^2, so the signal in this region is rather smooth. Finally, we predicted T_{xx}, T_{yy} and T_{zz} in the LORF in the inner region.

These LSC predicted gradients may be compared with true gradients (without errors) computed with EGM96, $n = 51 - 360$. Differences between these gradients are due to the 1 mGal error in the terrestrial data in combination with model errors such as spherical approximation in LSC and the use of a limited region in the computations (Table 2). The errors in the predicted gradients are small with almost zero mean and an error standard deviation of 1 mE. The predicted standard deviations from LSC are given between brackets and they closely resemble the true standard deviations σ (column 3 of Table 2). It is therefore feasible to calibrate the GOCE gradients using LSC and our simulated ground data. In addition, these data may be used for error assessment, see Figure 2.

Given the GOCE gradients, we would like to determine global calibration parameters, that is, calibration parameters valid for the whole time series of gravity gradients. However, it is not possible to discriminate between, e.g. a bias and 1 and 2 cpr effects in limited regions. In this study, we therefore estimated one bias per track, and could add, for example, one scale factor for the whole region. Also from

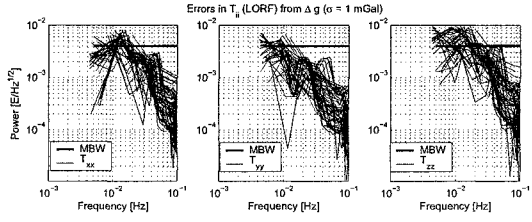


Fig. 2. PSD of the errors in the LSC predicted gravity gradients.

Table 3. Statistics of the errors in the predicted gradients after calibration: predicted minus true. Units are [mE]. T_{ii}: T_{xx} or T_{yy} or T_{zz} was used.

	T_{xx}		T_{yy}		T_{zz}	
observation	mean	σ	mean	σ	mean	σ
T_{ii}	-0.2	0.8	-1.2	0.8	1.9	1.0
T_{xx}, T_{yy}, T_{zz}	-0.1	0.2	-0.2	0.4	0.3	0.4
$T_{ii}, \Delta g$	-0.1	0.5	-0.3	0.1	0.6	0.2

the simulated GOCE gradients the contribution of EGM96, $n \leq 50$ was subtracted.

Gravity gradient biases were estimated for each track, for all three diagonal components, using LSC with parameters (Moritz, 1980). Different combinations of observations were used, with an assigned uncertainty of 5 mE for the gravity gradients and 1 mGal for the ground data. The bias errors per track (that is, the differences between the LSC predicted bias and the true bias) are small in all cases, that is, the bias error RMS is 1 mE or less. The 'true' bias is simply the mean simulated gravity gradient error per track, which is a lumped bias (sum of several instrumental errors).

The statistics of the errors in the calibrated gradients for the different cases are given in Table 3. The mean error of the calibrated gradients is almost zero for all cases, while the error standard deviation is at the mE-level. The error σ's have become smaller compared to the error in the observed gradients, also when only the gradients were used in the calibration. This is due to the fact that the LSC takes the signal correlation into account and smooths the data.

3 Validation

3.1 Band-limited kernels

In the operational phase of the GOCE mission there is a need to check the validity of the calibration. Here a method for monitoring and validation is proposed to be used during the measurement phases. The idea is to compare regional gravity and satellite gradiometer data from GOCE in a spectrally band-limited sense, where the band-limits need to be tuned to keep most of the gradiometer signal in the MBW. In case we isolate a bandpass filtered satellite gradiometer contribution and apply the same bandpass filter to modify the Stokes kernel, we are able to compare and possibly validate the results directly in a limited area in terms of geoid heights or disturbing potential, or even gravity or gravity gradients (cf. Arabelos and Tscherning, 1998; Koop et al., 2001). The initial focus will be on the potential at the Earth's surface or geoid heights, because the main interest is to analyze the effect of a possible build up of systematic errors or change in calibration characteristics of the accelerometers in the final end products. There is no need to consider the low frequency part of the signal because this is for the most part determined by SST, whereas the high frequency part is determined mainly by the ground data. For gradiometer monitoring purposes it is therefore sufficient to study a bandpass filtered signal. See also (Haagmans et al., 2003).

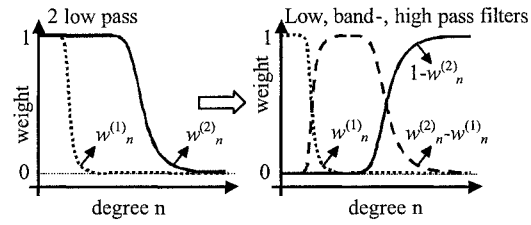


Fig. 3. Combination of two lowpass filters yields a lowpass, a bandpass, and a highpass filter.

A band-limited solution for the potential at the Earth's surface T_P^M, either from gravity anomalies Δg or from vertical gravity gradients of GOCE T_{rr}^G at satellite altitude is:

$$T_P^M = \frac{R}{4\pi} \int_\sigma S(\Psi, R) \Delta g_Q d\sigma \quad (2)$$

$$= \frac{r_G^2}{4\pi} \int_\sigma F(\Psi, r_G)(T_{rr}^G)_Q d\sigma \quad (3)$$

with R the average Earth radius, r_G the geocentric radius of the GOCE orbit, Ψ the spherical distance between points P and Q, and in which the kernels for surface gravity and vertical gravity gradients for GOCE are

$$S(\Psi, R) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} w_n^M P_n(\cos\Psi) \quad (4)$$

$$F(\Psi, r_G) = \sum_{n=2}^{\infty} \left(\frac{r_G}{R}\right)^{n+1} \frac{2n+1}{(n+1)(n+2)} \times w_n^M P_n(\cos\Psi) \quad (5)$$

with n the spherical harmonic degree, P_n Legendre function, and band pass filter $w_n^M = w_n^{(2)} - w_n^{(1)}$, where $w_n^{(1)}$ and $w_n^{(2)}$ are two lowpass filters (see Figure 3). Eq. (2) and (3) are the global solutions for gravity and satellite gradient data to obtain the medium wavelength part of the disturbing potential.

Next a regional solution will be derived starting from the global ones. For this purpose one needs to consider the following steps:

1. Divide the solutions for gravity and gravity gradients in inner and outer zone components;
2. Design adequate bandpass filter;
3. Derive the corresponding omission errors;
4. Estimate the truncation capsizes from omission error analysis;
5. Apply these parameters in regional solutions.

We can evaluate the integral components in Eq. (2) and (3) within a certain integration cap and omit the part outside this cap. Therefore we need to define spherical caps Ψ_0 and Ψ_1 for the integration of gravity and gravity gradient data by fixing one value for the omission error variance for geoid heights or the potential. An adequate choice of the truncation caps should imply that the regional contribution for any selected area solely based upon gravity or on

satellite gravity gradients has exactly the same truncation error, which means that the commission parts are the same for a 'deterministic case' with continuous signals without measurement errors, and the omission error parts as well. Based upon a degree-variance model c_n^{TR} for gravity anomalies, i.e. Tscherning/Rapp's model (Tscherning and Rapp, 1974), both geoid height omission error variances can be estimated (see Heck, 1979).

The spatial truncation caps at respectively Ψ_0, Ψ_1 of the 'modified' kernel functions S and F in Eq. (2) and (3) depend on the choice of the weights w_n^M, and on the admissible size of the truncation error variance. So, by analyzing the omission error variance for various capsizes given chosen weight functions a selection is done for a fixed variance. In case the omission error is evaluated at the Earth's surface, the downward continuation term is dominant, especially for higher degrees resulting in an unstable operator. For a meaningful practical application one has to create a bandpass filter that counteracts the spectral characteristics of the downward continuation operator $\frac{1}{(n+1)(n+2)}(\frac{r_G}{R})^{n+1}$. Ideally a spherical bandpass filter is desired that keeps the signal within the measurement bandwidth of the accelerometers. This means that the error correlation between the gradiometer data is drastically reduced allowing a limitation of the integration to a certain region only for space wise approaches (cf. ESA, 2000). Numerical tests show that filters can be designed such that the omission error stays below the required accuracy level, for example, 1 cm in terms of geoid heights (cf. Haagmans et al., 2003).

3.2 Satellite track crossovers

We compare measured gradients with other measured gradients in satellite track crossovers (XO). This technique has been applied very successfully in satellite altimetry to adjust (and calibrate) the observations. Albertella et al. (2000) have indicated that the GOCE gravity gradients V_{zz} may be compared in crossovers with sufficient accuracy. Crossovers allow to detect possible systematic temporal errors, to check the inherent accuracy of the gradiometer system and to give accuracy information, which may be helpful as input for the collocation.

The V_{zz} component is the most simple element of the gradient tensor for crossover comparison because it is approximately radially oriented (besides a small rotation about the cross-track axis, which has to be considered in the following computations), and therefore it could be compared easily in a real satellite track crossover. To compare the other in-line elements of the gravity tensor, V_{xx} and V_{yy}, the tracks have either to intersect orthogonally or one has to consider the different orientation of the gradiometer axes along each track by rotating the measurements or by applying a reduction. As a reduction will be necessary for the height differences (see below) and in order to avoid a quality loss of the measurements by the rotation, the reduction concept was chosen.

In reality one cannot just take the measured values and compare them, but the following steps are necessary. First,

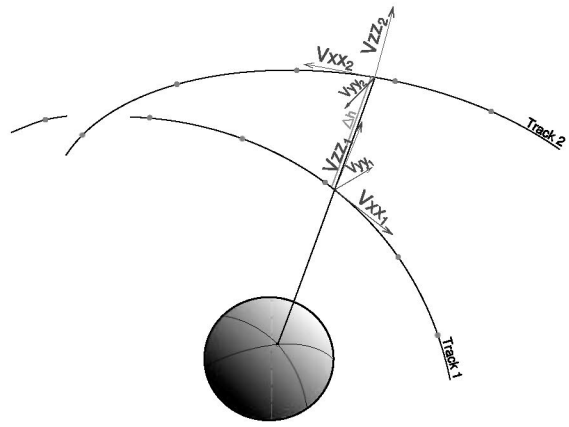


Fig. 4. XO geometry.

one has to identify the crossovers, which do not really exist three-dimensionally because of the orbit eccentricity. Therefore they are computed on a two-dimensional reference surface (e.g. a sphere), which defines also the projection direction. The computation has been performed numerically in the earth-pointing reference frame using again the IAG SC7 GOCE test data set. Each ground track crossover has been found by the intersection of two parabolas defined by the four sample points close to the specific crossover (see Figure 4). When the crossover has been found, the measured gradients, the satellite heights and orientations have been interpolated to this point. As we did not apply any special filter, the full gradient signal has to be considered. Therefore and because of the large distance between the samples (5 s corresponding to about 40 km) in the test data set, a linear interpolation of the gradients, orientations and heights failed. This led us to a quadratic approach for the interpolation using bivariate polynomials, see (Müller et al., 2003).

As already mentioned above, the orbit eccentricity (about $5 \cdot 10^{-3}$) causes height differences Δh_{12} up to 12 km in the crossovers, which leads to measurement differences up to 15 E in V_{zz}. That means, the gradients have to be reduced to the same reference height and, as stated in the second paragraph, to the same orientation. The so-called height and orientation reduction can be applied by

$$V_{ii}(h_1) = V_{ii}(h_2) - V_{iih}(h_1)\Delta h_{12} - V_{ii\alpha}(h_1)\Delta\alpha_{12}. \quad (6)$$

The derivative of the gradients in the height direction V_{iih} and the derivative with respect to the satellite orientations $V_{ii\alpha}$ is computed from the EGM96 model up to degree and order 300 by simple numeric differentiation together with the height (Δh_{12}) and orientation ($\Delta\alpha_{12}$) differences.

Using EGM96 in each calculation step as well as for the computation of the observations, a closed-loop study has been carried out. The resulting differences in the in-line components ΔV_{ii} are less than 0.6 mE (see Figiure 5), in agreement with the accuracy of the interpolation of the input parameters (XO position, satellite height and orientation differences and measured gradients); similar results are obtained with other

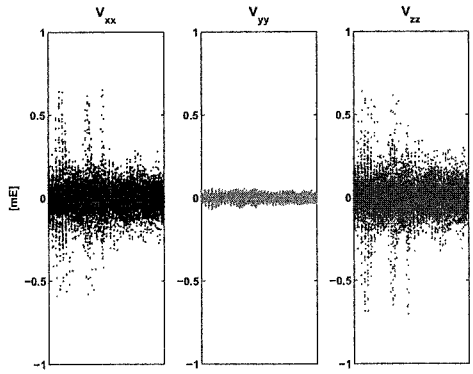


Fig. 5. Consistency check: gradient differences in XOs of IAG SC7 data.

models like OSU91A, too. In a further investigation, the coefficient variances of EGM96 have been taken into account. Their error propagation shows a Δh-dependent accuracy for the crossover reduction of (max) 2.5 mE for Δh's of 12 km. In the majority (over 65%) of the crossovers the reduction errors are below 1 mE (Jarecki and Müller, 2003). Considering a GOCE accuracy of about 2 mE, the used algorithms turn out to be sufficient for the internal validation of the gradients.

4 Conclusions and discussion

We discussed a number of methods for the calibration and validation of GOCE gravity gradients. The difference between calibration, including error assessment, and validation is that no corrections are applied to the data in the validation procedure.

We have shown that (existing) global gravity field models may be used for external calibration of the GOCE gradients. Although these models are not as accurate as the gravity field information we may derive from GOCE, they are expected to be more accurate at longer wavelengths below the MBW. However, if the GOCE gravity gradient scale factors, for example, are frequency independent, then we are also able to (indirectly) calibrate in the MBW with global models. A large part of the systematic errors at long wavelengths can be removed.

Terrestrial gravity data may be used for external calibration error assessment and validation (see also Denker, 2003). The number of regions, however, with high quality terrestrial data is limited and it may therefore not be possible to calibrate the full, global GOCE gradient data set. With LSC it is possible to accurately calibrate or validate the gradients at satellite level. In addition, the upward continued and converted terrestrial data may be used for error assessment. The validation of Level 2 data, such as the disturbing potential, is possible in a band-limited integral equation approach. Gravity gradients are downward continued and bandpass filtered. In a regional solution, with limited integration area, the disturbing potential can be computed which may be compared against a regionally derived disturbing potential from gravity anomalies, also bandpass filtered. The logical next step is to insert LS prediction into the integral equations.

Finally, we have studied one example of 'internal' validation, that is, we validated GOCE gravity gradients at satellite track crossovers. The method may benefit from the use of LSC, but this requires further study.

Acknowledgement. We acknowledge the review by Thomas Gruber which improved the paper.

References

Albertella, A., Migliaccio, F., Sansò, F., and Tscherning, C., Scientific data production quality assessment using local space-wise pre-processing, in *From Eötvös to mGal, Final Report*, edited by H. Sünkel, ESA/ESTEC contract no. 13392/98/NL/GD, 2000.

Arabelos, D. and Tscherning, C., Calibration of satellite gradiometer data aided by ground gravity data, *Journal of Geodesy*, *72*, 617–625, 1998.

Bouman, J. and Koop, R., Error assessment of GOCE SGG data using along track interpolation, *Advances in Geosciences*, *1*, 27–32, 2003a.

Bouman, J. and Koop, R., Geodetic methods for calibration of GRACE and GOCE, *Space Science Reviews*, *108*, 293–303, 2003b.

Denker, H., Computation of gravity gradients for Europe for calibration/validation of GOCE data, in *Gravity and Geoid 2002; 3rd Meeting of the IGGC*, edited by I. Tziavos, pp. 287–292, Ziti Editions, 2003.

ESA, Gravity Field and Steady-State Ocean Circulation Mission, Reports for mission selection; the four candidate earth explorer core missions, ESA SP-1233(1), 1999.

ESA, From Eötvös to mGal, Final report, ESA/ESTEC Contract No. 13392/98/NL/GD, 2000.

Haagmans, R., Prijatna, K., and Omang, O., An alternative concept for validation of GOCE gradiometry results based on regional gravity, in *Gravity and Geoid 2002; 3rd Meeting of the IGGC*, edited by I. Tziavos, pp. 281–286, Ziti Editions, 2003.

Heck, B., Zur lokalen Geoidbestimmung aus terrestrischen Messungen vertikaler Schweregradienten, Reihe C No. 259, Deutsche Geodätische Kommission, 1979.

Jarecki, F. and Müller, J., Validation of GOCE gradients using crossovers, in *GEOTECHNOLOGIEN: Observation of the System Earth from Space*, GEOTECHNOLOGIEN Science Report No. 3, 2003.

Knudsen, P., Estimation and modelling of the local empirical covariance function using gravity and satellite altimeter data, *Bulletin Géodésique*, *61*, 145–160, 1987.

Koop, R., Visser, P., and Tscherning, C., Aspects of GOCE calibration, in *International GOCE user workshop*, vol. WPP-188, ESA/ESTEC, 2001.

Koop, R., Bouman, J., Schrama, E., and Visser, P., Calibration and error assessment of GOCE data, in *Vistas for Geodesy in the New Millenium*, edited by J. Ádám and K.-P. Schwarz, vol. 125 of *International Association of Geodesy Symposia*, pp. 167–174, Springer, 2002.

Moritz, H., *Advanced physical geodesy*, Wichmann, 1980.

Müller, J., Jarecki, F., and Wolf, K., External calibration and validation of GOCE gradients, in *Gravity and Geoid 2002; 3rd Meeting of the IGGC*, edited by I. Tziavos, pp. 268–274, Ziti Editions, 2003.

Tscherning, C., A FORTRAN IV program for the determination of the anomalous potential using stepwise least squares collocation, Report No. 212, Department of Geodetic Science and Surveying, Ohio State University, 1974.

Tscherning, C., Computation of covariances of derivatives of the anomalous gravity potential in a rotated reference frame, *Manuscripta Geodetica*, *8*, 115–123, 1993.

Tscherning, C. and Rapp, R., Closed covariance expressions for gravity anomalies, geoid undulations, and deflections of the vertical implied by anomaly degree variance models, Report No. 208, Department of Geodetic Science and Surveying, Ohio State University, 1974.

Impact of space-borne gravity gradiometer instrument filter on observation error and gravity field recovery performance for GOCE

P.N.A.M. Visser

Delft Institute for Earth-Oriented Space Research, Delft University of Technology, Kluyverweg 1, 2629 HS, Delft, The Netherlands, Phone: +31 15 2782595, Fax: +31 15 2785322
email: Pieter.Visser@lr.tudelft.nl

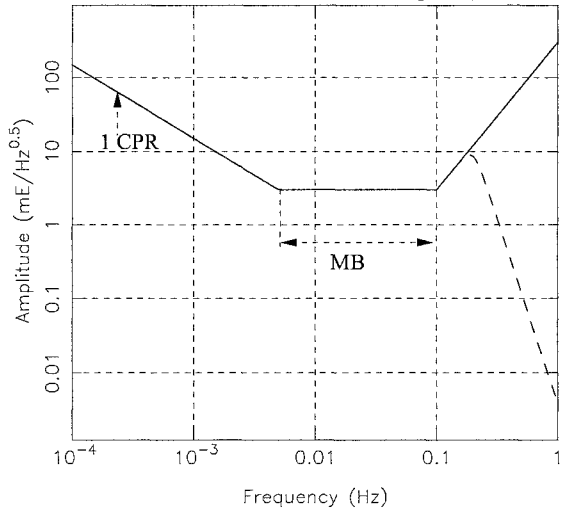
Abstract. The primary instrument on board of the future ESA gravity field observing satellite GOCE is a gravity gradiometer. The observations from this instrument will be passed through a 7th-order Butterworth filter in order to reduce high-frequency noise and aliasing due to the output rate of 1 Hz. This filtering will distort the gravity gradients that are to be measured, which can however be largely compensated by a time tag correction. An assessment has been made of the remaining observation error after this correction and its impact on gravity field recovery performance. It has been found that the filtering leads to a very small increase of the gravity field recovery error budget at the medium to short wavelengths after application of the optimal time tag correction. Only at the very long wavelengths a slight degradation might take place. However, it is anticipated that the GOCE gravity field recovery will be improved/dominated by GPS satellite-to-satellite tracking observations at these long wavelengths.

Key words. space-borne gravity gradiometer, Butterworth filter, time tag correction, gravity field recovery performance

1 Introduction

The Gravity field and steady-state Ocean Circulation Explorer (GOCE), the first European Space Agency (ESA) core Earth Explorer Mission (EEM), is foreseen to be launched in 2006/2007 (http://www.esa.int/export/esaLP/goce.html, accessed June 2003). A complement of instruments will enable the recovery of the global gravity field with an unprecedented resolution of better than 100 km and an accuracy of about 1 cm and 1 mgal in terms of geoid and gravity anomalies (ESA 1999) enhancing the capabilities of the CHAMP and GRACE missions (Reigber et al. 2002; Reigber et al. 2003; Tapley and Reigber 1999). This complement of instruments includes a gravity gradiometer, a Global Positioning System (GPS) receiver and a drag-free control system. The gradiometer consists of an orthogonal triad of three pairs of ultra-sensitive accelerometers. According to the defined specifications, the gradiometer must be able to observe the diagonal gravity gradient components with a precision of 3 mE/$\sqrt{Hz}$ (σ_n) in the measurement bandwidth (MB) of 0.005-0.1 Hz (Figure 1). Outside the MB, the noise increases proportional with the square of the frequency at the high frequencies, e.g. about 100 σ_n at 1 Hz, and inversely proportional at the lower frequencies, e.g. about 30 σ_n at 1 cycle per orbital revolution (cpr) or 1.8×10^{-4} Hz. Especially the power of the noise at the high frequencies builds up very fast. Therefore, the gradiometer instrument will be augmented by electronic circuits that effectively eliminate high-frequency noise. In addition, the observations that are included in the telemetry are anticipated to have a time interval of 1 sec (1 Hz) and this filtering will reduce aliasing of high-frequency (> 1 Hz) gravity gradient signals. The drawback of the filtering device is, however, that the gravity gradient signals will be distorted. An assessment of the severity of this distortion will be made by means of a simulated set of satellite gravity gradient observations. In a second step, attention will be paid to compensating this distortion by applying corrections. Finally, the impact of remaining gravity gradient observation

Fig. 1. Nominal noise profile for the diagonal components of the GOCE gravity gradiometer (solid line). The dashed line represents the noise that remains after application of a 7th-order Butterworth filter with a corner frequency of 0.2 Hz.



error after applying these corrections on gravity field recovery performance will be predicted.

2 Space-borne gradiometer

The primary instrument on board of GOCE is a gravity gradiometer consisting of three orthogonal pairs of accelerometers. Each accelerometer is capable of observing accelerations in two sensitive and one less-sensitive directions. The highest sensitivity is of the order of 10^{-12} m/s$^2/\sqrt{Hz}$ in the MB (0.005-0.1 Hz). The separation between the two accelerometers of one pair is about 0.5 m (ESA 1999). The three pairs of accelerometers are configured in such a way that all three diagonal gravity gradient components (Γ_{xx},Γ_{yy}, Γ_{zz}) and one off-diagonal gravity gradient component (Γ_{xz}) can be measured with high precision and the other two off-diagonal components (Γ_{xy},Γ_{yz}) with less sensitivity (use is made of a local vertical, local horizontal - LVLH - orthonormal reference frame, where x, y, z represent the along-track, cross-track and radial direction for a nominal circular orbit, respectively).

In order to prevent the high noise levels outside the MB, the gravity gradients are in the current design of the instrument passed through a $7th$-order Butterworth filter (priv. comm., R. Floberghagen, ESTEC, Noordwijk, The Netherlands, January 2003). The frequency response function $H_n(j\omega)$ of this filter satisfies the following property (Oppenheim and Schafer 1975):

$$|H_n(j\omega)|^2 = \frac{1}{1+(\frac{\omega}{\omega_c})^{2n}} \quad (1)$$

where, n is the order (equal to 7), ω the frequency (rad/s), and ω_c the cut-off frequency (rad/s). The pole locations s_k are (odd n, normalized ω_c=1):

$$s_k = e^{j\frac{2k}{n}\frac{\pi}{2}} \quad k = 1, 2, ..., n \quad (2)$$

where $j = \sqrt{-1}$.

For the GOCE gravity gradiometer, the cut-off frequency is equal to 0.2 Hz ($\omega_c = 0.4\pi$ rad/s), well beyond the MB. The frequency response in terms of magnitude and phase of the 7th-order Butterworth filter is displayed in Figure 2 and the reduction of the noise is included in Figure 1 (only visible at frequencies above 0.2 Hz). It can be seen that noise and signal amplitudes will be reduced very significantly at the high frequencies. However, at lower frequencies, also within the MB, significant phase shifts can be expected. Although high-frequency noise is reduced very efficiently, the question needs to be addressed how much the gravity gradients are distorted and how well this distortion can be modeled and/or corrected.

3 Satellite gravity gradients

In order to assess the impact of the filtering on the observed gravity gradients, a data set was simulated using a high degree and order reference gravity field model. First, unfiltered observations are described. Second, the changes in the observations due to the filter are analyzed. It is assumed that the gravity gradiometer axes are perfectly aligned with the axes of the LVLH reference frame.

3.1 Unfiltered observations

The analyses are based on one repeat period of about one month of simulated satellite gravity gradients. Use was made of a GOCE-type circular repeat orbit with an altitude of about 250 km and an inclination of 96 deg. The satellite completes 467 revolutions in a repeat period of 29 days. The EGM-96 gravity field model was used complete to degree and order 360 to represent the Earth's gravity field (Lemoine et al. 1997). Such a model has a spatial resolution of approximately 50 km or a smallest wavelength of 100 km. Assuming a satellite velocity of 7.8 km/s this corresponds with a highest frequency of 0.08 Hz, below the cut-off frequency of the Butterworth filter. As mentioned

Fig. 2. Bode diagram for 7th-order Butterworth filter. The figure on the left displays the amplitude reduction factor RF in dB or $20 \times^{10} log(RF)$ as a function of the frequency, the figure on the left the phase shift.

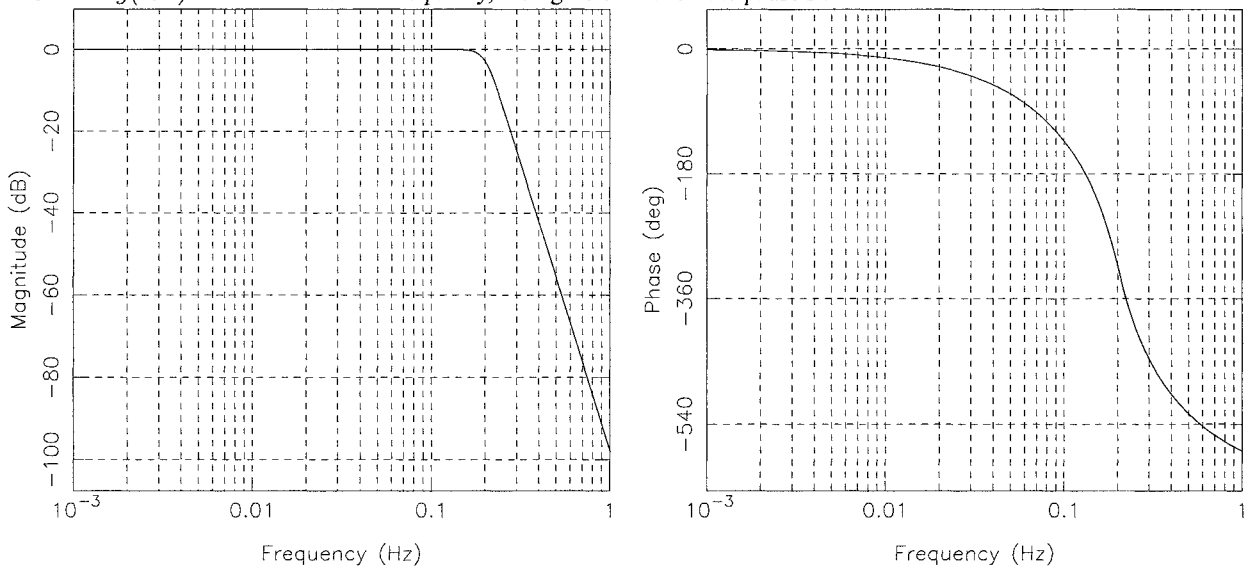
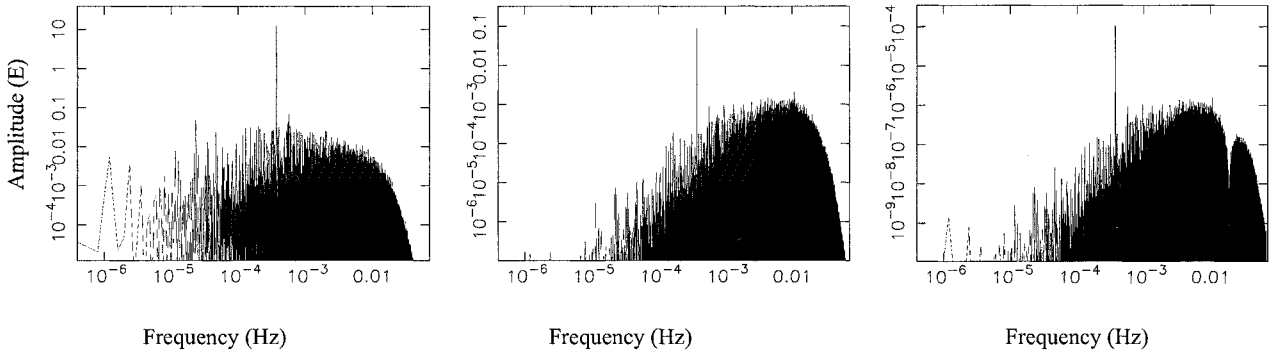


Fig. 3. Amplitude spectrum of signal (left), distortion (middle) and remaining observation error after application of a 3.58 sec time tag correction (right). Use was made of a $7th$-order Butterworth filter with a corner frequency of 0.2 Hz. The Earth's gravity field was represented by the EGM-96 gravity field model complete to degree and order 360.



before, the objective of this filter is to reduce high-frequency noise and at the same time minimize distortion of the gravity gradient signals themselves. It might be argued that a higher degree and order spherical harmonic expansion should have been used. However, GOCE aims at obtaining a global gravity field model with a resolution around 100 km (ESA 1999), above the 50 km resolution of a 360x360 spherical harmonic expansion. In addition, when using the Rapp'79 gravity anomaly degree variance model (Rapp 1979), the total estimated power of the diagonal radial gravity gradient for degrees above 360 is equal to 0.002 mE. It is therefore fair to assume that the results will not change significantly when a higher degree and order spherical harmonic expansion is used.

For a circular repeat orbit, the gravity gradients can be represented by a Fourier series making use of the Kaula expansion of the spherical harmonic gravity potential in Kepler elements (Kaula 1966). In the simulations, it will be assumed that the GOCE orbit is perfectly circular. In reality, the satellite will fly in a near-circular orbit with an eccentricity of less than 0.0045 leading to a small maximum pitch angle of 0.26 deg. Therefore, it is fair to assume that representative results are obtained when making use of a perfectly circular orbit.

The diagonal satellite gravity gradient term (Γ_{zz}) is the dominant component and will therefore be taken to serve as example. This component has

Fig. 4. Typical pass of filtered and unfiltered Γ_{zz} satellite gravity gradient observations.

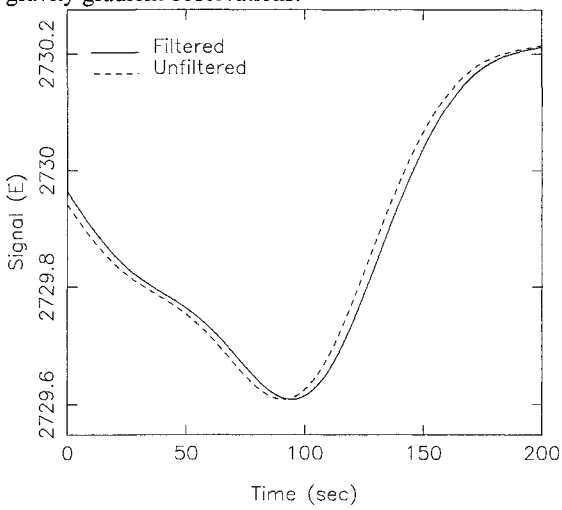
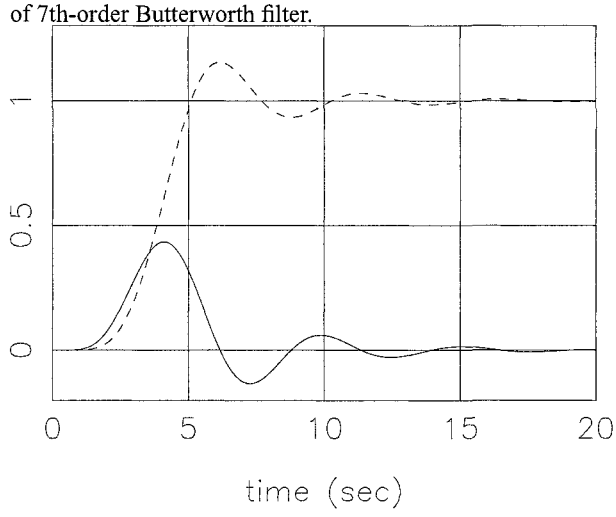


Fig. 5. Impulse (solid line) and step (dashed line) responses of 7th-order Butterworth filter.



a mean value of 2738.61 E (10^{-9} s^{-2}) and a Root-Mean-Square (RMS) about mean of 8.63 E.

The spectrum for Γ_{zz} is displayed in Figure 3 (left). A clear dominant peak can be distinguished at 2 cpr (cycle per orbital revolution), caused by the J2 term of the gravity field.

3.2 Filtered observations

The frequency response of the 7th-order Butterworth filter was applied to the frequency spectrum of the unfiltered satellite gravity gradient observations. First, the time series of gravity gradients was converted to a frequency spectrum by a fast Fourier transform. Second, the frequency spectrum of the filtered observations was obtained by applying amplitude attenuation factors and phase shifts according to Figure 2. The resulting distortion in the frequency domain, being the difference between the frequency spectra of the unfiltered and filtered observations, is displayed in Figure 3 (middle). Again, a clear dominant peak can be distinguished at 2 cpr. After converting the frequency spectra to the time domain, it can be shown that the filtering can be largely represented by a time shift of the observations (Figure 4). The Butterworth filter seems to cause a delay of about 3.6 sec, which is equivalent to about 28 km along the satellite flight path. This time shift can be largely explained by analyzing the pattern of the phase shift of the 7th-order Butterworth filter (Figure 2, right) in combination with the frequency spectrum of the signal (Figure 3). The signal is dominant at frequencies below 0.01 Hz. The phase shift can be approximated quite well by a straight line for frequencies below this value. In combination with a constant rotation rate of the satellite around the Earth, this slope can be translated to a time shift. In addition, this time shift can also be predicted to a good extent by analyzing the impulse response of the 7th-order Butterworth filter for a corner frequency of 0.2 Hz (Figure 5). The filtered signal can also be obtained by a convolution of the impulse response and the unfiltered observations in the time domain. It can be seen that the peak of the impulse response is close to 4 sec.

4 Correcting the observations

Figure 4 suggests that the effect of the Butterworth filter on the gravity gradient observations can be largely compensated by a time tag correction. Several values for the time tag correction were applied to the filtered gravity gradient frequency spectra and compared with the unfiltered spectra (Table 1). An optimal time tag correction was found equal to 3.58 sec, close to the value obtained after visual inspection of Figure 4. The remaining distortion (Figure 3, right) is around 0.08 mE compared to 76.2 mE for the non-corrected filtered gravity gradients. The time tag correction is thus very efficient in removing the effect of the filtering for observed gravity gradients. However, still a small dominant 2 cpr peak with an amplitude of about 0.1 mE can be observed.

5 Global gravity field recovery

After applying the time tag correction, small gravity gradient errors remain (Figure 3, right). The

Table 1. Comparison between unfiltered and time shifted filtered satellite gravity gradients for different time tag corrections (all 6 components)

Time tag (sec)	RMS of fit (mE)
0.0	76.155
3.0	12.275
3.50	1.625
3.58	**0.080**
3.6	0.505
3.8	4.765
4.0	9.025
5.0	30.323

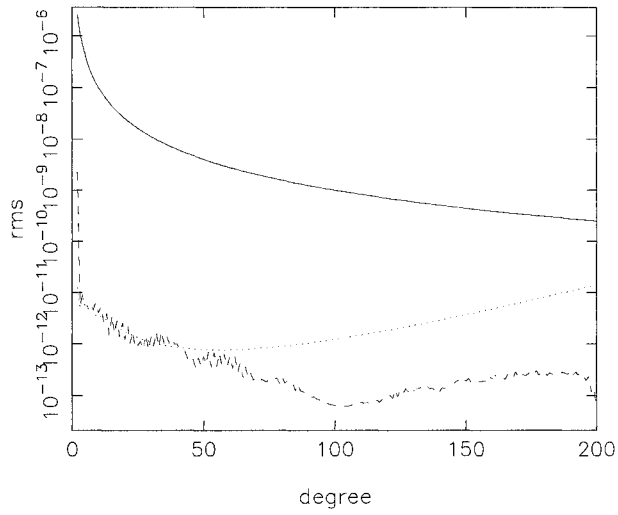


Fig. 6. Global gravity field recovery formal error estimates. The rms of the spherical harmonic coefficient errors is displayed as a function of the degree (solid line: Kaula's rule of thumb, dotted line: predicted GOCE performance (Visser et al. 2002), dashed line: after time tag correction).

Fourier spectrum of this remaining error is assumed to be representative for all diagonal satellite gravity gradient components. This Fourier spectrum is propagated to gravity field recovery error by means of a covariance analysis (ESA 1999; Colombo 1984) assuming that all three diagonal gravity gradient components are observed (Figure 6). It has to be noted that the observation period was taken equal to one repeat period or 29 days. The GOCE gravity field solution will nominally be based on two observation periods of 6 months, thus 12 months in total, with a hibernation period in between (ESA 1999). It might thus be argued that the gravity recovery error will be reduced by a factor of about $\sqrt{12}$. However, this is only true when recovery errors of successive months are uncorrelated. It is fair to anticipate that the spectrum of the remaining gravity gradient error (Figure 3, right) will be similar for successive months. Therefore, no reducting factor was applied to take into account the 12 month observation period. The predicted error displayed in Figure 6 can thus be considered to represent a worst case. As indicated above, the GOCE gravity field solution will be based on 12 months of GPS satellite-to-satellite tracking and satellite gravity gradiometer observations. Nominally, use will be made of the three diagonal gravity gradient components (although one off-diagonal term might become available with high precision as well and will then be included in the gravity solution). The gravity field recovery performance for this nominal mission is included in Figure 6 (Visser et al. 2002). At the time the prediction for this nominal mission was made, it was not known that the gravity gradients would be passed through a filter and its effect was not taken into account. At the medium to high degrees (70-200) the predicted gravity error for the nominal mission is, however, about an order of magnitude above the predicted error due to remaining filter correction errors. Therefore, the gravity field recovery error budget will not increase significantly for these degrees. At the low degrees, the predicted gravity recovery error for the nominal mission and due to the filtering correction errors have about the same order of magnitude (apart from a peak at degree 2 caused by the 2 cpr gravity gradient correction error). However, at the low degrees the final GOCE gravity solution will be dominated/improved significantly by the GPS tracking observations.

6 Conclusions

The satellite gravity gradient observations that will be observed by the GOCE gradiometer will be affected due to the application of a $7th$-order Butterworth filter. However, the resulting distortion, about 76 mE for the Γ_{zz} component, can almost be entirely compensated by an appropriate time tag correction and reduced to about 0.080 mE. The remaining observation error does not lead to a significant deterioration of the anticipated GOCE gravity field recovery performance except for the very low degree terms. However, especially at the low degrees, the GPS SST observations will improve/dominate the GOCE gravity field solution, thereby further reducing the impact of the remaining gravity gradient observation errors.

References

Colombo, O. L. (1984), *The Global Mapping of Gravity With Two Satellites*, vol. 7, no. 3, Netherlands Geodetic Commission, Publications on Geodesy, New Series.

ESA (1999), Gravity Field and Steady-State Ocean Circulation Mission, *Reports for Mission Selection, The Four Candidate Earth Explorer Core Missions, SP-1233(1)*, European Space Agency, July 1999.
Kaula, W. M. (1966), *Theory of Satellite Geodesy*, Blaisdell Publishing Co, Waltham, Massachusetts.
Lemoine, F. G., *et al.* (1997), The development of the NASA GSFC and NIMA Joint Geopotential Model, in *International Association of Geodesy Symposia, Gravity, Geoid and Marine Geodesy*, vol. 117, pp. 461–469, Springer-Verlag, Berlin.
Oppenheim, A.V., and R.W. Schafer (1975), *Digital Signal Processing*, Englewood Cliffs, NJ: Prentice Hall.
Rapp, R.H. (1979), Potential coefficient and anomaly degree variance modeling revisited, Report no. 292, Department of geodetic science, OSU, Ohio, Columbus.
Reigber, Ch., G. Balmino, P. Schwintzer, R. Biancale, A. Bode, J.-M. Lemoine, R. Koenig, S. Loyer, H. Neumayer, J.-C. Marty, F. Barthelmes, F. Perosanz, and S.Y. Zhu (2002), A high quality global gravity field model from CHAMP GPS tracking data and Accelerometry (EIGEN-1S), *Geophys. Res. Lett.*, *29*(14), 1–4.
Reigber, Ch., P. Schwintzer, K.-H. Neumauer, F. Barthelmes, R. König, Ch. Förste, G. Balmino, R. Biancale, J.-M. Lemoine, S. Loyer, S. Bruinsma, F. Perosanz, and T. Fayard (2003), Earth Gravity Field Model EIGEN-2, *Adv. Space Res.*, *31*(8), 1883–1888.
Tapley, B.D., and C. Reigber (1999), GRACE: a satellite-to-satellite tracking geopotential mapping mission, in *Bollettino di Geofisica Teorica ed Applicata, Vol. 40, No. 3-4, Sep.-Dec. 1999, Proceedings of the 2nd Joint Meeting of the International Gravity and the International Geoid Commission, Trieste 7-12 Sept. 1998, ISSN 0006-6729*, edited by I. Marson and H. Sünkel, p. 291.
Visser, P.N.A.M., R. Rummel, G. Balmino, H. Sünkel, J. Johannessen, M. Aguirre, P.L. Woodworth, C. Le Provost, C.C. Tscherning, and R. Sabadini (2002), The European Earth Explorer Mission GOCE: Impact for the Geosciences, in *Ice Sheets, Sea Level and the Dynamic Earth, Geodynamics Series 29, American Geophysical Union*, edited by J. Mitrovica and L.L.A. Vermeersen, pp. 95–107.

Testing frame-transformation, gridding and filtering of GOCE gradiometer data by Least-Squares Collocation using simulated data

C.C.Tscherning, Department of Geophysics, University of Copenhagen, Juliane Maries Vej 30, DK 2100 Copenhagen Ø, Denmark., e-mail: cct@gfy.ku.dk

Abstract: Least-squares collocation (LSC) has been used to test frame transformation, filtering and gridding of GOCE gradiometer data. Initially the function of the GRAVSOFT program GEOCOL was checked by using noise-free data with 5 s sampling provided by IAG along a 1 month realistic orbit. Data in a 2° x 2° area was used. For these data it was verified that 0.1° frame transformation and gridding in an interior 1° x 1° area of one quantity from the same quantity could be done with an error below 5 % of the signal standard deviation. If several components of the gravity gradient were used, the error decreased slightly.

Noise with standard deviations of 0.050 and 0.150 Eötvös Units ($E=10^{-9}/s^2$), respectively, and a 1 s correlation distance was generated and added to the data. For the 0.050 E noise the result of the gridding was as for the data without noise, when 1 s sampling was used. For the 0.150 E data, the use of totally 2 months of simulated data and all 3 diagonal components of the gradient matrix were necessary in order to obtain a result below 10 % of the signal standard deviation. Further improvements are possible if data from a larger area or from the planned second measurement phase of GOCE are used.

1. Introduction.

It will most likely not be possible to use least-squares collocation (LSC) in its most general form for the modeling of the gravity field of the Earth using GOCE (ESA, 1999) satellite gradiometer data. In its general form LSC requires the solution of systems of equations with as many unknowns as observations.

Regional solutions may however be constructed, and such solutions may form the basis for the creation of datasets of gridded values at satellite altitude. Certain types of gridded data (e.g. the second order radial derivative, $d^2T/dr^2=T_{rr}$, where T is the anomalous potential) may be used as the basis for the application of the method of Fast Spherical Collocation, see Sansò and Tscherning, (2002).

In order to test the use of LSC, the GRAVSOFT program GEOCOL (Tscherning, 1975, Tscherning et al., 1994) has been used with test data available from the IAG Special Commission 7. These test data have been generated for a period of one month with a 5 s sampling, along a realistic GOCE orbit. It includes all the 6 second order derivatives of the gravity potential, V, given in a local reference frame aligned with the velocity vector of the satellite and the radius vector. No roll of the satellite was included. The data were generated using EGM96 (Lemoine et al., 1998) to degree 300. The data set did not include any noise. However, noise has been simulated and test was carried out using the noisy data. The data set has furthermore been densified time-wise (to 1 s) and space-wise corresponding to 2 months of observations in order to test the influence of the sampling.

In the following we will describe the results of the testing. Not only did the testing reveal some embarrassing programming errors, but it did also give better insight into which errors could be expected when doing frame transformation and gridding. In section 2 the test using data without noise, and in section 3 the results of the gridding using data with added noise are described.

2. Testing the LSC prediction procedure with data without noise.

As a reference potential EGM96 to degree 24 was used. This has the effect that the fields become more statistically homogeneous. This is not very important for the gridding, but very important for the interpretation of the error estimates. Furthermore the data decorrelates to a certain extend.

As a covariance function we used a function defined using the EGM96 error-degree variances e_i^2 from degree 2 to 24 scaled with a factor of a=0.5, the depth to the Bjerhammar-sphere of 1061 m and the variance of the gravity anomalies at the mean Earth surface of

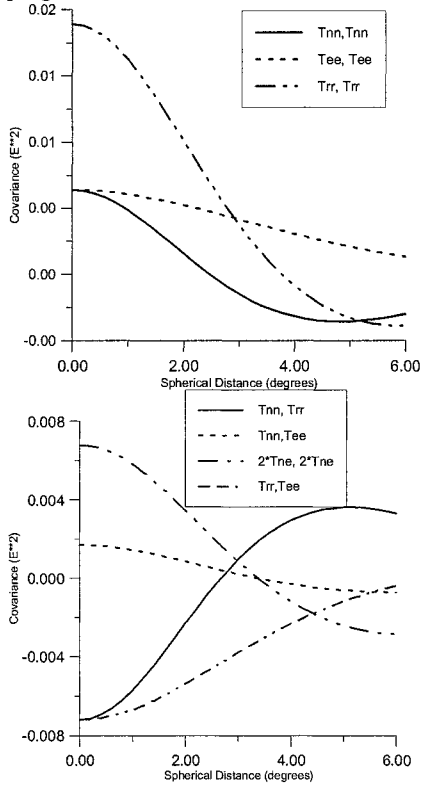
408.13 mgal^2. A scale factor on the degree-variances from 25 to infinity was used equal to A= 98.79 mgal^2/R^2, where R is the mean Earth radius. Hence the analytic covariance expression for the anomalous potential is with R_B the Bjerhammar sphere radius

$$\mathrm{cov}(P,Q) = \mathrm{cov}(r,r',\psi) =$$

$$\sum_{i=2}^{24} a * e_i^2 \left(\frac{R_B^2}{rr'}\right)^{i+1} P_i(\cos\psi) +$$

$$\sum_{i=25}^{\infty} \frac{A}{(i-1)(i-2)(i+4)} \left(\frac{R_B^2}{rr'}\right)^{i+1} P_i(\cos\psi)$$

P_i are the Legendre polynomials, r and r' are the radial distances from the origin of P, Q, respectively, and ψ is the spherical distance between the points. Covariances of the gravity gradient components (Fig. 1) may be evaluated as described in Tscherning (1976, 1993). Subroutines COVAX, COVBX and COVCX used for these computations are included in the program GEOCOL.



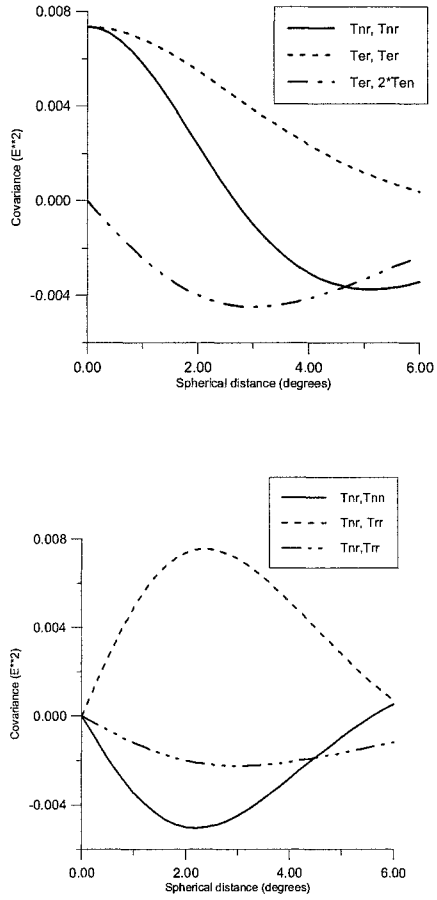


Figure 1. Covariance functions of gravity gradients at 255 km altitude in an East (e), North (n) and radial (r) frame as a function of spherical distance. An azimuth equal to zero was used. This makes covariances of all combinations of gravity gradients not shown identical zero.

Two test areas were selected. One bounded at latitudes –1° and 1° and longitude –1° and 1°. This area included 29 points. The other area was bounded by latitude 81 and 83 degrees and longitude 45 to 50 degrees and included 294 points. However this area was only used initially because results similar to the first mentioned area were obtained.

The individual gravity gradients $T_{kl}(Q_n)$, with error correlation σ_{mn}, were used to predict the other gradients, $T_{ij}(P)$

$$\tilde{T}_{ij}(P) = \left\{\mathrm{cov}(T_{kl}(Q_m), T_{ij}(P))\right\}^T \cdot \left\{\mathrm{cov}(T_{kl}(Q_m), T_{kl}(Q_n)) + \sigma_{mn}\right\}^{-1} \cdot \left\{T_{kl}(Q_n)\right\}$$

as well as the components of the gravity vector given in an inertial, Cartesian frame (also computed from EGM96 to degree 300). The evaluation of the covariances require that the functionals associated with the observations (second order derivatives in local cartesian frames) are applied on the fundamental covariance expression given above. We will use two frames: the instrument frame (along-track: x, direction of radius vector: z, y: orthogonal on x,z) and the e: East, n: North and r: radially up frame in a geocentric CTRS. The **LSC** error estimates were also computed, see (Moritz, 1980, eq. 14-42).

Below in Table 1, the results from the first test area are presented. The points in which the values are predicted are the same points, which are associated with the observations. Hence, the test is a kind of consistency test, and it was used to detect errors in the subroutine COVCX. It furthermore shows how well/bad a pointwise frame-transformation may be done.

The results are given as the differences and standard deviations of "observed" minus calculated and the mean value of the error of prediction calculated also by **LSC**. Units are mE. In the calculations an uncorrelated data noise of 5 mE was used, i.e. 25 mE^2 was added to the diagonal of the normal equations.

The table shows that there is a very good consistency between the different quantities in the sense that the analytically computed error of prediction (in bold) is large when the error is large and small when it is small. Components of the gravity gradient, which are spatially close (T_{zz} and T_{rr}) are predicted with a small error (3 mE), while components which are spatially different (T_{xx} and T_{ee}) have a large error (64 mE) compared to the signal variance (71 mE).

It also illustrates the well known fact, that we using only local data obtain biased results when we predict one functional from another. See for example the third row, where the prediction of T_{rr} has a standard deviation of only 5 mE with a bias of 39 mE when predicted from T_{xx}.

Similar results were obtained in the area close to the Pole, where the velocity vector direction varies much. It was furthermore throughout the computations verified that the Laplace-equation was fulfilled both in the East/North/Up (ENU) frame and in the instrument frame. This shows that the program works correctly at least for this type of quantities. However, before this conclusion could be reached a number of errors were detected and corrected.

Pre-dicted	T_{ee}	$2T_{en}$	T_{er}	T_{nn}	T_{nr}	T_{rr}
LSC St.de.	***071***	***082***	***089***	***074***	***089***	***127***
Data						
T_{xx}	-34	006	-18	001	046	039
	005	007	022	002	006	005
LSC	**064**	**014**	**060**	**003**	**027**	**063**
$2T_{xy}$	-10	000	-15	-08	077	019
	002	001	006	005	012	007
LSC	**044**	**003**	**037**	**044**	**033**	**088**
T_{yz}	001	-10	000	017	008	-18
	007	013	001	009	006	011
LSC	**025**	**028**	**003**	**035**	**016**	**046**
T_{yy}	000	003	-18	005	132	-05
	001	006	003	031	010	031
LSC	**003**	**016**	**031**	**063**	**061**	**062**
T_{xz}	-10	-20	-01	022	000	-12
	013	024	007	006	001	008
LSC	**034**	**029**	**015**	**022**	**003**	**044**
T_{zz}	-10	-45	-18	010	044	000
	010	026	008	010	008	001
LSC	**030**	**057**	**028**	**029**	**026**	**003**
T_{xx}, T_{yy}, T_{zz}	000	-03	000	000	-06	000
	001	002	041	000	009	001
LSC	**002**	**008**	**021**	**002**	**024**	**002**

Table 1. Results of predicting (e,n,r) gravity gradient components from instrument (xyz) gravity gradients. EGM96 to degree 24 subtracted. The first column gives the kind of data used for prediction. The x, y, z refers to derivatives of T along-track, cross-track and up. The column heading of each of the following columns shows the type of quantity predicted. The second row give calculated variances calculated from the covariance model. Mean and standard deviation of the predicted quantity is found in the upper cell and below in bold the mean value of the prediction error computed using GEOCOL. Units now and later are mE.

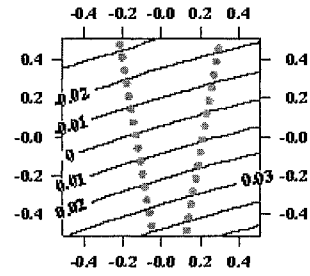


Figure 2. Area used for test of gridding. Contour lines of T_{rr} in E based on EGM96 degree 25-300 coefficients. The dots are simulated GOCE measurement points.

Subsequently data in a 0.1° grid bounded by -0.5° and 0.5° in latitude and longitude (Fig. 2) were generated in an East, North, radially up frame. The excellent results, which are obtained when a quantity is spatially close to the one used for the prediction and when several components are used, are given in Table 2.

Predi -cted Data	T_{ee}	$2T_{en}$	T_{er}	T_{nn}	T_{nr}	T_{rr}
xx	032 004	007 003	-017 010	001 000	042 002	040 002
LSC	**060**	**011**	**057**	**003**	**027**	**059**
2*xy	-011 001	000 000	-014 003	009 001	069 005	020 002
LSC	***043***	***003***	***030***	***043***	***030***	***017***
xz	010 007	016 009	000 003	021 003	000 001	-011 004
LSC	**030**	**028**	**013**	**021**	**002**	**042**
yy	000 000	-002 001	003 001	007 002	-003 006	-007 002
LSC	**003**	**016**	**028**	**059**	**058**	**059**
yz	000 004	007 004	000 000	019 005	005 002	-020 006
LSC	**024**	**026**	**003**	**032**	**017**	**042**
zz	-009 005	-041 011	-018 003	009 005	040 003	000 000
LSC	**027**	**053**	**025**	**027**	**025**	**002**
xx,yy ,zz	000 000	002 001	-009 001	001 000	038 001	000 000
LSC	**002**	**007**	**021**	**002**	**020**	**002**

Table 2. The table gives results of predictions of data in a grid with spacing 0.1° totally 121 points. The first row gives the mean and standard deviation of the differences "observed" minus predicted. The second row gives the LSC computed error estimate. Units: mE. Note that the instrument x-axis in this area points close to North or South in an Earth-Fixed frame. This is the reason that results of predicting T_{nn} from T_{xx} or T_{ee} from T_{yy} are excellent. See Fig. 2.

Again the results are striking. The mean values are sometimes large, indicating that we are using a too limited area of "observations". Again the noise-free observations in the calculations were supposed to have an associated noise standard deviation of 5 mE, i.e. 25 mE^2 were added in the diagonal of the normal equations. This mainly affects the LSC error-estimate. Note that the computed error-estimate is large when the bias is large, as it should be.

3. Gridding with noisy data.

An accepted model for the (total) gradiometer noise is given in the Granada report (ESA, 1999, Fig.8.2). This model has been used by (Regguzoni, 2002) for the generation of noise, which had a very large correlation distance. The elimination of this kind of noise requires that data are distributed over a much larger area than studied here, see (Tscherning and Arabelos, 2003) or (Bouman et al. 2003).

Using a procedure similar to the one used by (Regguzoni, 2002) pseudo-random noise was generated with a short correlation distance, so that the noise was uncorrelated after a 2 s time difference.

The generated noise was added to the "observations" used in section 2. These data *have a 5 s time-difference; hence they will be* regarded as uncorrelated. The variance of 0.003027 E^2 was added to the diagonal of the normal equations. The results are presented in Table 3.

Predi -cted Data	T_{ee}	$2T_{en}$	T_{er}	T_{nn}	T_{nr}	T_{rr}
xx	-012 002	029 005	-011 005	003 002	073 006	009 004
LSC	**068**	**061**	**071**	**016**	**046**	**071**
2*xy	-041 007	013 011	-044 008	071 027	190 015	-030 033
LSC	**059**	**018**	**062**	**063**	**062**	**111**
xz	004 007	015 010	001 004	028 004	004 006	-032 011
LSC	**066**	**052**	**064**	**035**	**017**	**081**
yy	-002 001	-060 005	000 007	023 004	002 014	-021 005
LSC	**015**	**065**	**048**	**069**	**071**	**071**
yz	-015 010	-012 013	001 005	065 005	008 014	-050 013
LSC	**037**	**050**	**017**	**065**	**068**	**082**
zz	-004 004	-029 004	-019 005	-005 013	123 002	009 015
	038	069	043	038	040	018

Table 3. Differences between "observed" and predicted data in a grid. Noisy data with a noise standard deviation of 55 mE was used. First row gives the type of predicted quantities and the first column the kind of data used. Units mE. By a mistake the noise added to the 2*xy component was not multiplied by 2.

In the table we note that not only have the predictions become worse, but also the estimated prediction error from LSC - as it should be. There are still some large biases due to the small area used for the "observations".

The observation and the prediction grid points are the same as used in Section 2.

It is encouraging to see that rather small errors of prediction can be expected using T_{zz} and predicting the same kind of data in a grid. The estimated error is 18 mE. Hence for 5 s means the error will be around 9 mE. Then we are not far from the 5 mE required to reach the goals of the GOCE mission.

A dataset with 1 s sampling was generated and noise added. This improved the results considerably, so when 3 diagonal components were used a "perfect" filtering was obtained, see Table 4.

Predicted Data	T_{ee}	$2T_{en}$	T_{er}	T_{nn}	T_{nr}	T_{rr}
xx, yy, zz	007 004	-002 003	-029 006	-001 002	054 003	-006 005
LSC	**006**	**031**	**029**	**006**	**029**	**006**
	xx	xy*2	xz	yy	yz	zz
xx, yy, zz	-003 005	001 010	001 061	004 007	-012 030	-001 010
LSC	**009**	**033**	**033**	**009**	**033**	**010**

Table 4. Prediction (row 2 and 3) and filtering (row 6 and 7) using 1 s T_{xx}, T_{yy} and T_{zz} data with noise standard deviation of 55 mE. Note the very good filtering at the data. The noise is reduced to 20 %.

Here a noise standard deviation slightly smaller equal to 0.116 E was used. The results using the 1 s - 1-month dataset were unsatisfactory, and a dataset corresponding to a 2-month period was generated. The results are given in Table 5.

Predicted Data	T_{ee}	$2T_{en}$	T_{er}	T_{nn}	T_{nr}	T_{uu}
xx,yy zz,	-004 002	-028 004	-029 004	-007 001	058 005	011 002
LSC	**008**	**041**	**031**	**008**	**032**	**009**
.	T_{xx}	T_{xy}	T_{xz}	T_{yy}	T_{yz}	T_{zz}
xx,yy zz,	-009 006	026 014	-001 068	-003 005	-009 032	011 004
LSC	**012**	**043**	**036**	**012**	**035**	**013**

Table 5. Gridding and filtering of data with 116 mE standard deviation noise added. T_{xx}, T_{yy} and T_{zz} used. Totally 720 observations, corresponding to 2 months of data, 1 s sampling. Rows 2 and 3 show the results of the gridding and rows 6 and 7 the filtering. The noise is reduced to (12 mE and 13 mE), i.e. 10 %

We see that having the double number of data enables us to obtain very satisfactory results when the data have a twice larger noise standard deviation as shown in Table 5. However a second measurement-phase of 2 months (as planned for GOCE) should give results of both the gridding and the filtering which are very satisfactory, i.e. like working with data with a 5 mE noise standard deviation.

4. Conclusion.

We should keep in mind that the error of prediction varies from area to area due to the variation in signal variance. We have here used a small area with a rather smooth gravity field. However, globally predicted T_{rr} give similar results, see (Tscherning and Arabelos, 2003) where equiangular 20° x 20° areas were used. This is probably due to the spatial closeness of the T_{zz} and the T_{rr} vector components.

LSC may (as demonstrated above) be used to combine different components, which will reduce the error, or equivalently improve the filtering of the noise. This naturally requires that the noise is uncorrelated between the gravity gradient components.

The tests carried out are both a test of the GEOCOL program as well as of the possible results to be expected when gridding noisy GOCE gradiometer data in order to use Fast Spherical Collocation.

Acknowledgement: Thanks to IAG SC7 for providing the test data.

References:

Bouman, J., R.Koop, R.Haagmans, J. Mueller, N.Sneuw, C.C.Tscherning and P.Visser: Calibration and Validation of GOCE Gravity Gradients. Poster prepared for IUGG/IAG General Assembly, Sapporo, 2003.

ESA: Gravity Field and Steady-State Ocean Circulation Mission, ESA SP-1233, 1999.

Lemoine, F.G., S.C. Kenyon, J.K. Factor, R.G. Trimmer, N.K. Pavlis, D.S. Chinn, C.M. Cox, S.M. Klosko, S.B. Luthcke, M.H. Torrence, Y.M. Wang, R.G. Williamson, E.C. Pavlis, R.H. Rapp, and T.R. Olson, The Development of the Joint NASA GSFC and the National Imagery and Mapping Agency

(NIMA) Geopotential Model EGM96, NASA/TP-1998-206861, Goddard Space Flight Center, Greenbelt, MD, July, 1998.

Moritz, H,: Advanced Physical Geodesy. H. Wichmann Verlag, Karlsruhe, 1980.

Reguzzoni, M.: From the time-wise to the space-wise GOCE observables. Como, 2002.

Sanso', F. and C.C.Tscherning: Fast spherical collocation: A General Implementation. IAG Symposia, Vol. 125, pp. 131-137, Springer Verlag, 2002.

Tscherning, C.C.: A FORTRAN IV Program for the Determination of the Anomalous Potential Using Stepwise Least Squares Collocation. Reports of the Department of Geodetic Science No. 212, The Ohio State University, Columbus, Ohio, 1974.

Tscherning, C.C.: Covariance Expressions for Second and Lower Order Derivatives of the Anomalous Potential. Reports of the Department of Geodetic Science No. 225, The Ohio State University, Columbus, Ohio, 1976.

Tscherning, C.C.: Computation of covariances of derivatives of the anomalous gravity potential in a rotated reference frame. Manuscripta Geodaetica, Vol. 18, no. 3, pp. 115-123, 1993.

Tscherning, C.C. and D.Arabelos: Computation of a geopotential model from GOCE data using Fast Spherical collocation - a simulation study. Prepared for IUGG/IAG General Assembly, G03, Sapporo, 2003.

Tscherning, C.C., R.Forsberg and P.Knudsen: The GRAVSOFT package for geoid determination. Proc. 1. Continental Workshop on the Geoid in Europe, Prague, May 1992, pp. 327-334, Research Institute of Geodesy, Topography and Cartography, Prague, 1992.

PROCESSING OF CHAMP ACC AND RSO DATA IN THE EARTH GRAVITY FIELD DETERMINATION

J. Luo
School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan, China

J. Sh. Ning
Laboratory of Geospace Environment and Geodesy, Wuhan University, 129 Luoyu Road, Wuhan, China

Abstract. The high frequency accelerometer can provide non-gravitational perturbation force acceleration acting on CHAMP satellite. In CHAMP standard science product level-2 data, the acceleration, star camera, thruster firing and some corrections data are available. They play important roles in the determination of the earth gravity field. In this paper, the level-2 data (Day of year 1-100, 2002) are processed. Before applying the corrections, the curves of mean acceleration per day are very rough. After corrections, the curves show clear trend. On the other hand, the precise orbit of CHAMP is necessary for the determination of the earth gravity field. Now, the rapid science orbit data are provided by two organizations, GFZ and JPL. The data of a randomly chosen period (Day of year 267-269, 2002) from the two organizations are compared. The results of the comparison show that the difference between the two RSO data is notable, especially at the end part of orbit files. After getting rid of the end part of each file, there is no obvious system bias exists between the two RSO data and the magnitude of the difference keeps only about 0.1 m. Finally, some conclusions are drawn and some suggestions are given.

Keywords. CHAMP, Accelerometer, RSO

1 Introduction

The CHAMP (CHAllenging Minisatellite Payload) satellite, designed by GFZ (GeoForschungsZentrum Potsdam), was launched in middle of 2000. This project focuses on the study of earth gravity field, atmosphere and magnetic field. More detailed information may be accessed via the GFZ web site http://www.gfz-potsdam.de. The CHAMP onboard accelerometer data (ACC) can be regarded as the precise measurement of non-gravitational acceleration acting on the satellite to remove non-gravitational effect. The RSO (Rapid Science Orbit) can be used as the measurement of the earth gravity signal for the determination of earth gravity field and orbit refining. In this paper, the CHAMP ACC data are processed and the results are analyzed. On the other hand, the CHAMP RSO data provided by GFZ (Schmidt R., Baustett G. König R. and Reigber Ch., 2002) and JPL (Jet Propulsion Lab) are compared and some useful conclusions are drawn.

2 Installation of Accelerometer and Reference System

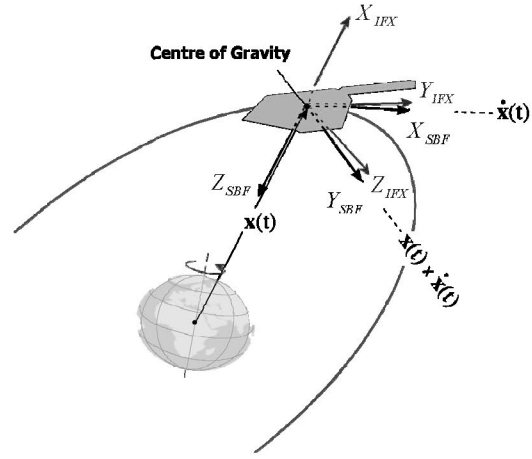


Fig.1 Reference Systems for CHAMP: X_{IFX} - anti-parallel to Z_{SBF}, Y_{IFX} - parallel to X_{SBF}, Z_{IFX} - parallel to Y_{SBF} (Modified from Schäfer Ch., 2000)

The ACC data of CHAMP need be transformed from Accelerometer Instrument Fixed System (IFX) to other various reference systems. The relationships among these reference systems are shown in Fig. 1. The IFX data can be transformed to Spacecraft Body Fixed System (SBF) with the satellite design parameters. And the data based on SBF can be transformed to Conventional Inertial Reference System (CIS) with attitude data (star camera data).

If the data based on Conventional Terrestrial Reference System (CTS) is required, the data based on CIS can be transformed to CTS (McCarthy D., 1996). For the requirement of POD(Precise Orbit Determination) and PGD (Precise earth Gravity field Determination), the CHAMP onboard accelerometer sensor is positioned precisely in the center of gravity of the CHAMP Satellite and the deviation is less than 2mm (GFZ, 2001). The exact axes orientations of accelerometer in space are measured by means of two star camera sensors. And the accelerometer stays in a temperature stability of better than 1 K in order to minimize thermal effects on bias and scale factor.

3 Non-gravitational Acceleration

Non-gravitational acceleration is commonly measured by onboard accelerometer. The ACC measurement $\ddot{\vec{R}}_A$ can be denoted with eq. (3.1). (Cappellari J. O., Velez C. D. and Fuchs A. J., 1976)

$$\ddot{\vec{R}}_A = \ddot{\vec{R}}_D + \ddot{\vec{R}}_{SR} + \ddot{\vec{R}}_E + \ddot{\vec{R}}_T + \ddot{\vec{R}}_{TH} \quad (3.1)$$

With:

$\ddot{\vec{R}}_D$ Atmosphere drag

$\ddot{\vec{R}}_{SR}$ Solar radiation pressure

$\ddot{\vec{R}}_E$ Earth radiation pressure

$\ddot{\vec{R}}_T$ Thermal of satellite

$\ddot{\vec{R}}_{TH}$ Thrust force

then the perturbation equation can be written as eq. (3.2).

$$\ddot{\vec{R}} = \ddot{\vec{R}}_{PM} + \ddot{\vec{R}}_{NS} + \ddot{\vec{R}}_{OE} + \ddot{\vec{R}}_A \quad (3.2)$$

With:

$\ddot{\vec{R}}_{PM}$ N-body perturbation

$\ddot{\vec{R}}_{NS}$ Earth non-spherical perturbation

$\ddot{\vec{R}}_{OE}$ Ocean tide and Solid Earth tide, etc

Because of the deviation of accelerometer center and CHAMP satellite mass center, the bias of accelerometer, the accelerometer scale factor variation, etc., the direct observations of onboard accelerometer cannot fully replace the components of right hand side of eq. (3.1). And the direct observations will be translated and corrected with eq. (3.3).

$$\ddot{\vec{R}}_A = Q(K\vec{A} + \vec{B}) \quad (3.3)$$

With:

Q Rotation matrix from IFX to integral system (eg. CIS)

K Scale factor matrix of accelerometer

$\vec{A}$ Observation of ACC based on IFX

$\vec{B}$ Bias correction vector

In this paper, Q is denote as eq. (3.4).

$$Q = Q_C Q_B Q_A \quad (3.4)$$

With:

Q_A Rotation matrix from IFX to SBF

Q_B Rotation matrix from SBF to CIS

Q_C Rotation matrix form CIS to custom system (optional)

And the Q_B is calculated with eq. .

$$Q_B = f(\theta,\ \phi,\ \psi) \quad (3.5)$$

With:

$\theta,\ \phi,\ \psi$ Three Euler angles

The Euler angles, which rotate the body about the three axes, are particularly popular for attitude determination and controlling. Although the Euler angles are visually helpful for understanding the rotational motion of satellite (yaw, pitch and roll), there is a disadvantage in the Euler angles, which is known as a singularity. The Euler angles are singular when the second Euler angle is 0° and 180°. So the quaternion representation is widely applied to the satellite PAD (Precise Attitude Determination) and control.

$$q = \begin{bmatrix} \vec{q} \\ q_4 \end{bmatrix} = [q_1, q_2, q_3, q_4]^T \quad (3.6)$$

The inverse of q is

$$q^{-1} \equiv \begin{bmatrix} -\vec{q} \\ q_4 \end{bmatrix} \quad (3.7)$$

$$C(q) = \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1q_2 + q_3q_4) & 2(q_1q_3 - q_2q_4) \\ 2(q_1q_2 - q_3q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2q_3 + q_1q_4) \\ 2(q_1q_3 + q_2q_4) & 2(q_2q_3 - q_1q_4) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{bmatrix} \quad (3.8)$$

The quaternions can be used to construct 3×3 rotation matrix, which have 12 different sets for the rotation matrix. And there is a unique combination of quaternions for the rotation matrix such as eq.

(3.8). And the attitude information can be translated from quaternions to Eurler angles or vice versa (Bae S. and Schutz B. E., 2001). In this paper, the quaternions are applied to the CHAMP acceleration data processing.

4 Result analyses

4.1 Day of year 1-100, 2002 ACC analysis

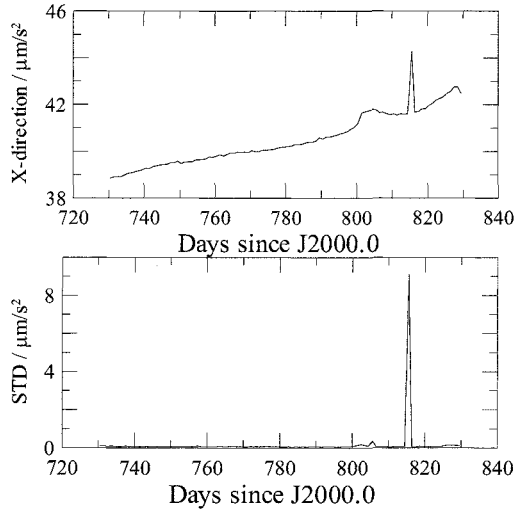


Fig.2 X-direction mean acceleration and STD per day

The standard science products are made available for the general co-investigators' user community via the ISDC in GFZ. Hereinto, the Orbit and Gravity (OG) level-2 product contains the preprocessed accelerometer observations (in 0.1 Hz), star camera data, thruster firing data and model correction data (Förste Ch. 2002, König R. 2001). In the model correction data, the CNES correction is only for X-direction. CHAMP satellite is a LEO satellite with about 450km altitude. And the atmosphere drag model, earth radiation model, etc., with high precision are usually unavailable. Moreover, the earth gravity determination needs to remove the non-gravitational acceleration with adequate precision. So the accelerometer data of CHAMP plays an important role in gravity field determination to replace all non-gravitational accelerations acting on CHAMP satellite. The effective accelerometer data processing needs to be combined with attitude data, which is provided by the star camera sensor. And they can be obtained from level-2 data.

Various schemes are adopted in the processing. The analyses results are shown in Fig. 2-7.

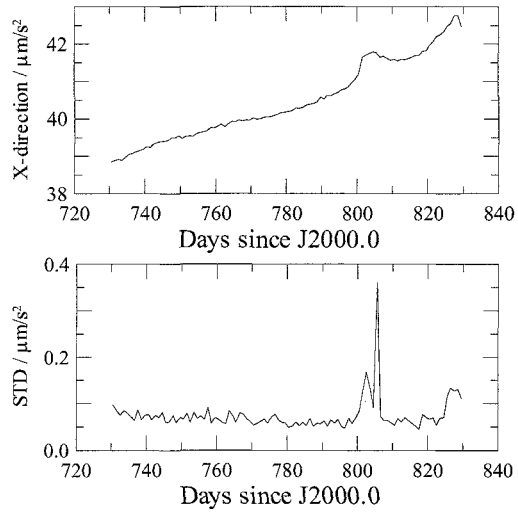


Fig.3 X-direction mean acceleration and STD per day, after getting rid of 815 day since J2000.0

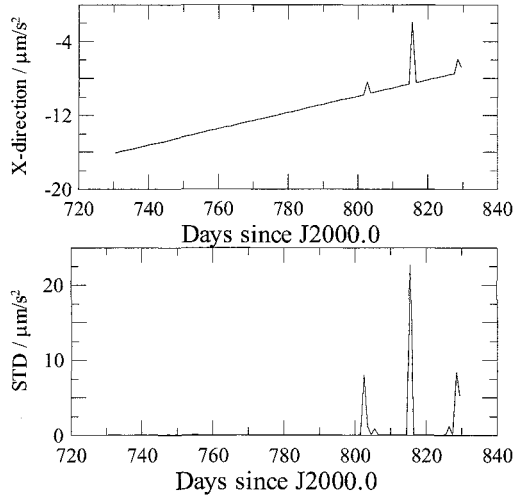


Fig.4 X-direction mean acceleration and STD per day, after correction

Before applying correction, which is provided by CNES, curves of mean acceleration per day are very rough (see Fig.2 and Fig.3). After the corrections are added, the trend lines become very smooth and straight except for a few anomalies (see Fig. 4 and Fig. 5). Sometimes, the trend lines show obvious spike (e.g. 815 day since J2000.0) and some small crests (e.g. 803 & 829 day since J2000.0). Also in these days the STD curves show larger value than normal status. They may be caused by the instability of the accelerometer. And such unstable status may last for one or several days. To prevent the importing of large errors, these data should be wiped off before calculating gravity field without

additional reference information. See Fig.4, there are still some small singulars in day 802, 803, 805, 826, 828 and 829 except the data of day 815. After removing them, the acceleration shows good performance in X-direction, see Fig. 5. On the other hand, the acceleration acting on the CHAMP satellite has a linear increasing trend. It may be caused by the residual bias even the CNES correction applied.

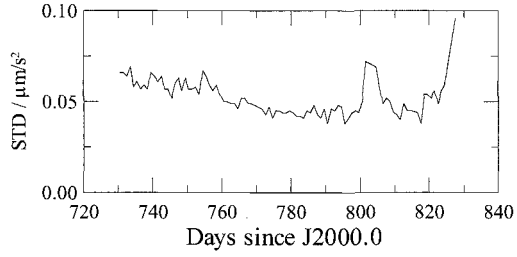


Fig. 5 X-direction STD per day, after getting rid of 802, 803, 805, 826, 828 and 829 day since J2000.0

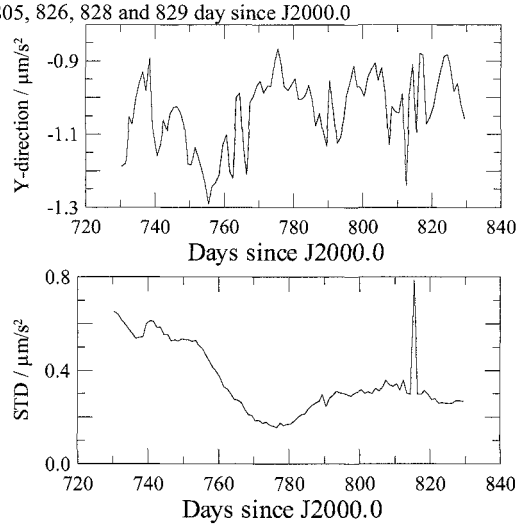


Fig. 6 Y-direction mean acceleration and STD per day

The correction of CNES has notable effect on the X-direction. The mean acceleration of 100 days changes from about 40 $\mu m/s^2$ to -12 $\mu m/s^2$, the acceleration direction turns upside down. At the same time, the trend curves show that there may still be residual "shift" existing in the accelerometer data, which may be caused by the changing of the seasons or the shift of accelerometer or both of them. The effect of the "shift" cannot be ignored, and how to deal with it will be an important part of further work.

Fig. 6 and Fig. 7 show the variation of acceleration in Y- and Z- directions in IFX. They show obvious anomaly in 815 day since J2000.0, the same as in X-direction. And we can see that the accelerations in Y- and Z-directions are smaller than that in X-direction with one magnitude level.

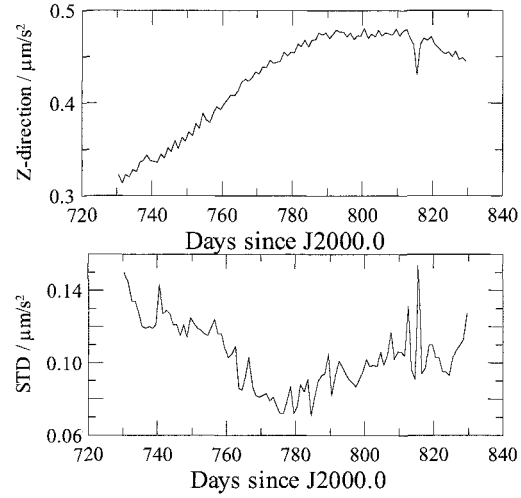


Fig. 7 Z-direction mean acceleration and STD per day

4.2 Day of year 267-269, 2002 RSO comparison analysis

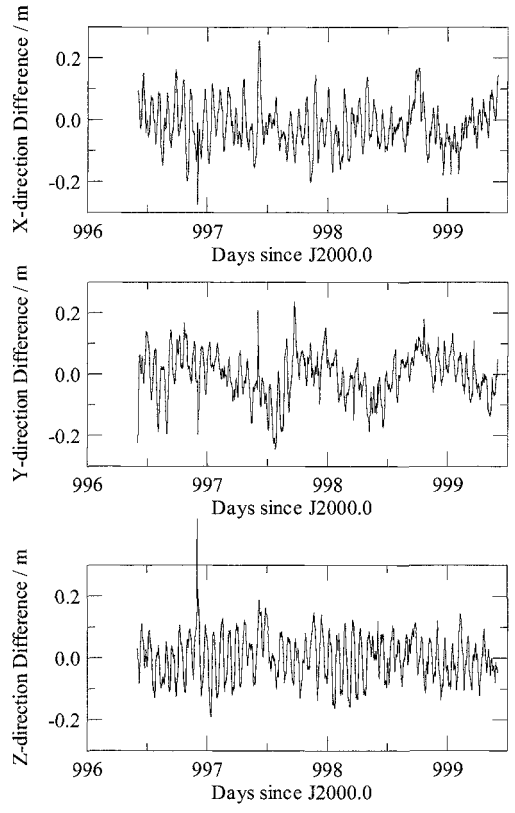


Fig. 8 Difference in X, Y and Z-direction (CTS) between GFZ and JPL RSO Data

To analyze the CHAMP orbit determination level, CHAMP RSO data of 3 days are randomly selected. These data are supplied by two organizations, GFZ and JPL. Because the two RSO data have different sampling rate, 0.0333Hz for GFZ and 0.0167Hz for JPL, the rate is uniformed in 0.0167Hz to avoid importing interpolation errors.

Fig. 8 shows the comparison between the two RSO data after getting rid of the tail parts of both RSO files. Because the difference between the two RSO data is 1m or even several meters in the end part of RSO files, which is unacceptable in data processing. When these RSO data are employed in the gravity field determination, the tail part of every RSO file should be removed as suggested. There is no obvious system bias exists between the two RSO data. And the magnitude of the difference keeps about 0.1 m. But as shown in Fig. 8, there is an unexpected deviation larger than 0.4m in Z-direction in 996.9 day since J2000.0. It may be caused by abnormal value in one or both RSO data.

5 Conclusion

After above analysis, some conclusions can be drawn:

- CHAMP onboard accelerometer has good performance, but the preprocessing of the data is very nesserary. The correction supplied by CNES has remarkable effect. After the correction, the every day mean acceleration has smooth and straight movement trend. In X-direction, the values change from positive to negative during the peroid Day of year 1-100, 2002.
- The CHAMP onboard accelerometer operates in unstable status sometimes. It may last for one day or several days. Its aftereffect is notable and the data need to be processed carefully.
- The non-gravitational accelerations acting on the CHAMP satellite are not in balance in different directions. The acceleration in X-direction is larger than those in Y- and Z-direction of one magnitude level.
- The acceleration acting on the CHAMP satellite has a linear increasing trend. It may be caused by that the accelerometer measurement contains residual bias even after applying the CNES correction.
- RSO data supplied by GFZ and JPL have about 0.1m difference, and this difference does not contain system error. When the RSO data are imported in processing, it is suggested that one but not the combination of the two RSO data be used.
- At the end part of every RSO data of GFZ's and JPL's, there is large difference of 1m or even several meters. These data need to be removed before further computation. Besides the end part of RSO data anomaly, some little anomaly need to be effectively dealt with.

6 Acknowledgement

Thanks to CHAMP ISDC of GFZ. Almost all source data are provided by ISDC. Thanks to Dr. Shi Chuang, who is of great help for our studying. And Thanks to Prof. Tscherning C. C's poster at Internet web site. And This work is supported by National Natural Science Foundation of CHINA (NSFC) project (40234039, 40174004), the key project of Chinese Ministry of Education and The Key Laboratory of Geospace Environment and Geodesy, Ministry of Education, China (LOGEG) fund project (02-09-04).

References

Bae S. and Schutz B. E., Geoscience Laser Altimeter System (GLAS) Precision Attitude Determination, Center for Space Research, The University of Texas at Austin, 2001.2

Cappellari J. O., Velez C. D. and Fuchs A. J., Mathematical Theory of Goddard Trajectory Determination System (N76-24291-24302), Goddard Space Flight Center, 1976.4

Förste Ch., Format Description:The CHAMP Data Format. GFZ-Potsdam, 2002.3

GFZ, Announcement of Opportunity for CHAMP. GFZ-Potsdam, 2001.5

König R., Format Description:The CHAMP Orbit Format CHORB. GFZ-Potsdam, 2001.6

McCarthy D., IERS Convention (1996). USA: U.S. Naval Observatory, 1996.7

Schäfer Ch., The Sensitivity Analysis of GPS-tracked Satellite Missions (Case Study CHAMP): Ph. D. Dissertation, The University of Stuttgart, 2000.10

Schmidt R., Baustett G. König R. and Reigber Ch., Orbit Predictions for CHAMP – Development and Status. First CHAMP Science Meeting, GeoForschungsZentrum, Potsdam, Germany, Jan 22 - 25, 2002

One year of time-variable CHAMP-only gravity field models using kinematic orbits

Nico Sneeuw
University of Calgary, Department of Geomatics Engineering, sneeuw@ucalgary.ca

Christian Gerlach, Lóránt Földváry, Thomas Gruber, Thomas Peters, Reiner Rummel, Dražen Švehla
Technische Universität München, Institut für Astronomische und Physikalische Geodäsie

Abstract. A full year of CHAMP gravity field solutions has been calculated using the energy integral approach. The monthly solutions in the time frame 03.2002–02.2003 were based solely on kinematic orbits from CHAMP GPS orbit tracking and accelerometry. These kinematic orbits have not been contaminated by a priori gravity field information.

Recovery of medium-wavelength time-variable gravity signal from CHAMP is expected to be at the edge of feasibility. This expectation is validated by comparison to an SLR solution of seasonal gravity variations and by comparison to oceanographic and meteorological models. It is shown that the error level of monthly CHAMP solutions is insufficient for revealing these time variations. Orbit decay—and consequently ground track variation—is a main contributor to this effect.

Keywords. CHAMP, Jacobi integral, time variable gravity field

1 Introduction

Gravity field determination from CHAMP data using the energy balance approach is successfully demonstrated, e.g. (Han et al., 2002) or (Gerlach et al., 2003c). First attempts of recovering time-variations in the gravity field from CHAMP data were published by Sneeuw et al. (2003) and Cheng et al. (2003). The former made use of the energy balance approach too.

These studies made clear that time-variable gravity recovery from CHAMP will be at the edge of feasibility. CHAMP's sensitivity is too low to reveal the full spectrum of seasonal and sub-seasonal variability. With proper care, perhaps the lowest spherical harmonic degrees can be revealed. In the aforementioned studies, two main obstacles were still present: *i)* Insufficient length of CHAMP time series. At least one full year of high quality orbits and accelerometry should be available. *ii)* Use of reduced-dynamic orbits. In (Gerlach et al., 2003a) it is demonstrated that gravity solutions from (reduced-) dynamic orbits are biased towards the a priori gravity field.

Recently, a full year of kinematic CHAMP orbits were published, (Švehla & Rothacher, 2003a), that strictly avoid the dependence on prior gravity information. Based on this data set a stationary CHAMP-only gravity field was determined by Gerlach et al. (2003a), denoted TUM-1S. Validation showed TUM-1S to be of equal or better quality than EIGEN-2, (Reigber et al., 2003; Gerlach et al., 2003b).

This orbital data set would overcome the two aforementioned obstacles. It is optimally suited to test CHAMP's sensitivity to temporal gravity changes. The next section will briefly describe the methodology and data used. After that the gravity recovery results are presented in spatial and spectral domain. As will be shown, the results are disappointing. The final section will focus on the role of orbit height and ground-track variation.

2 Data and methodology

Energy balance approach. Using the energy integral—or Jacobi integral—is a well documented and proven method for gravity field recovery from orbit data by now, e.g. (Han et al., 2002; Gerlach et al., 2003c). Its main merits are: *i)* simplicity, *ii)* separation of orbit and gravity field information (and corresponding processing), and *iii)* avoidance of initial state problems. It relies, though, on precise orbit determination (POD), preferably as a long and uninterrupted time-series of orbit positions and velocities. In particular the resulting energy profile is sensitive to velocity errors, cf. (Visser et al., 2003).

The energy integral (per unit mass) which is the basis for gravity field determination, reads:

$$T + c = E_{\mathrm{kin}} - U - Z - \int \boldsymbol{f} \cdot \mathrm{d}\boldsymbol{r} \qquad (1)$$

At the left we have the unknown disturbing potential T, up till an unknown constant c. All terms at the right are determined from CHAMP data:

- $E_{\mathrm{kin}} = \frac{1}{2}\boldsymbol{v}\cdot\boldsymbol{v}$, the kinetic energy, requires orbit velocities $\boldsymbol{v}$,
- U, the normal gravitational potential, requires satellite positions $\boldsymbol{r}$,

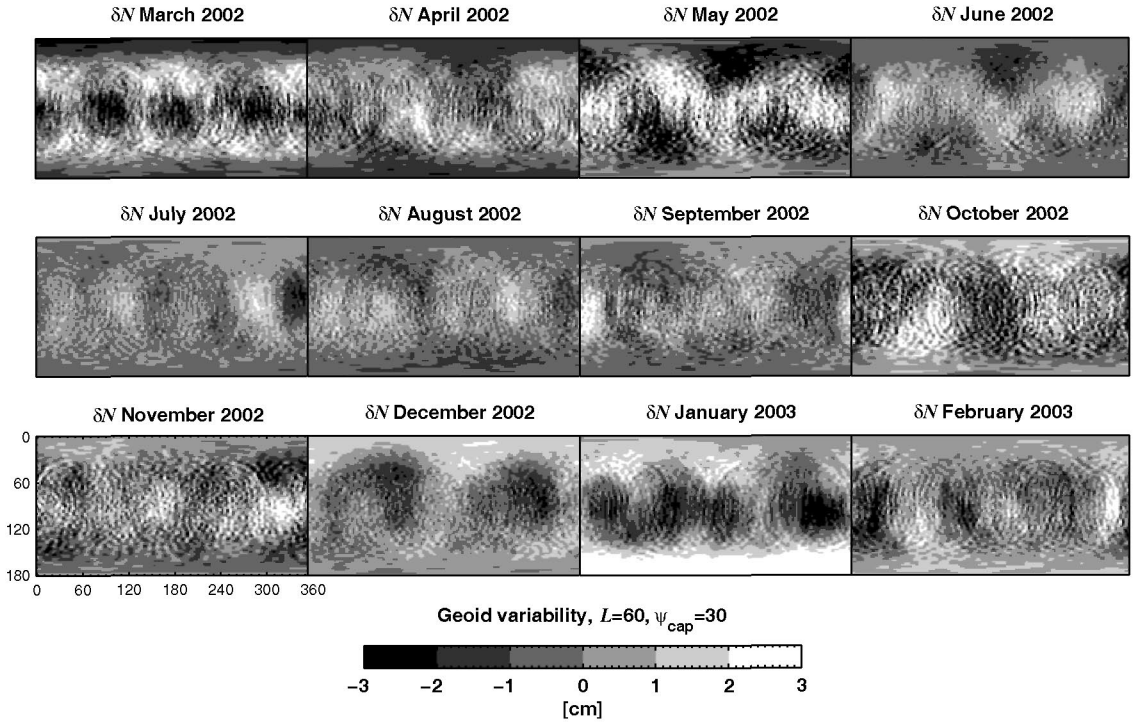


Fig. 1. Monthly CHAMP solutions, in terms of geoid residuals, smoothed with spherical cap of 30° radius.

- $Z = \frac{1}{2}(\boldsymbol{\omega}\times\boldsymbol{r})\cdot(\boldsymbol{\omega}\times\boldsymbol{r})$, the centrifugal potential at the satellite's location,
- $\int \boldsymbol{f}\cdot \mathrm{d}\boldsymbol{r}$ is the dissipated energy, which is an integral of CHAMP's accelerometer data $\boldsymbol{f}$ along the orbit,

Boldface symbols denote vectors. Note that the dissipative energy mainly integrates the along-track force.

Data sets. One full year of CHAMP GPS data (10 March 2002 – 9 February 2003) has been processed according to the methodology of Švehla & Rothacher (2003b). Zero-difference ambiguities and epochwise CHAMP positions and clock parameters were estimated from zero-difference carrier phase using CODE derived GPS satellite clock and earth rotation parameters. As a result, a full year of high quality 3D orbit positions has become available with a reported accuracy of 1–3 cm, (Švehla & Rothacher, 2003a).

Kinematic orbit determination does not provide orbit velocities. They were determined in a separate post-processing step from the orbit positions using smoothing splines, cf. (Gerlach et al., 2003a). The sampling rate for both $\boldsymbol{r}$ and $\boldsymbol{v}$ is 30 s. The accuracy of the velocity is estimated to be 0.1 mm/s or less

CHAMP accelerometry was used to evaluate the dissipative energy integral in (1). However, due to mismodelled scale factor and in particular to accelerometer bias, this integral potential would show a strong drift. Thus, the resulting disturbing potential is high-pass filtered above $1.2 \cdot 10^{-4}$ Hz.

Moreover, ancillary data (e.g. attitude quaternions) and models (e.g. tides) are used in the evaluation of the disturbing potential. Any known time-variable gravity contributions can be dealt with similarly in the dissipation integral.

Spherical harmonic analysis. The resulting time series of disturbing potential T is reduced to a constant altitude, e.g. mean orbital altitude, and subsequently gridded onto a torus, cf. (Sneeuw, 2001). Spherical harmonic analysis is performed using inclination functions. The radial continuation and the gridding induce errors that are iteratively eliminated.

Pre-mission error analyses and current CHAMP solutions indicate that one can expect a *degree of resolution* around $L = 60$. Then a Nyquist-type rule of thumb says that we need at least twice as much orbital revolutions, i.e. 120. With the given orbital revolution rate—around 16 revolutions per day—we would need approximately one week of data. Previous experience, (Sneeuw et al., 2003), has shown that high redundancy reduces sampling problems.

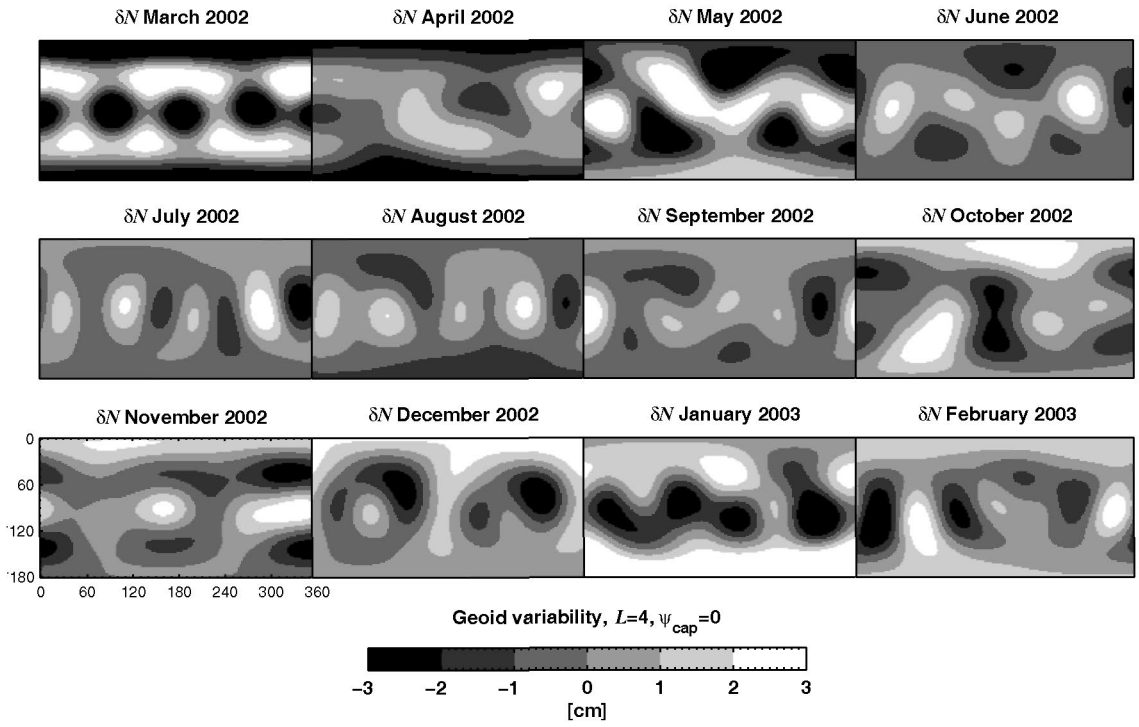


Fig. 2. Monthly CHAMP solutions, in terms of geoid residuals, up to maximum degree $L = 4$.

Thus, the data have been separated into monthly batches, leading to monthly gravity field solutions up to $L = 60$. The solution *March 2002* actually stands for the time frame 10.03.2002–9.04.2002, and so on. Each monthly solution is considered to be stationary. The time-series of all solutions provides the time-variable gravity field. Naturally, all shorter period phenomena will alias into the individual solutions. It is believed, though, that with the present accuracy of CHAMP data, this effect is insignificant.

3 Results

Spatial domain. Although the monthly gravity fields were solved up to $L = 60$, sensitivity analyses and previous studies have demonstrated that only the very long wavelengths of time-variable signal may be revealed by CHAMP. Thus the results have to be filtered. We choose two approaches here: *i)* spatial filtering with a smoothing cap of 30° radius, and *ii)* spectral filtering by spherical harmonic degree cut-off at $L = 4$. Both filter techniques represent extremes. The chosen cap size and cut-off degree are somewhat arbitrary. However, they serve the purpose of focusing on the long wavelenghts, where time-variable signal is expected.

Figure 1 shows the monthly geoid solutions, corrected for a mean geoid, with the 30°-cap smoothing. The variations are in the ± 3 cm range. A lot of high frequency content is still present in the filtered geoid residuals because of the wide spectral transfer of Pellinen coefficients.

Figure 2 avoids this problem by rigorously low-pass filtering up till $L = 4$. The resulting monthly residuals are very smooth indeed. The range of variations, roughly ± 3 cm, compares well to reported and expected values of geoid change, due to atmospheric, oceanographic or hydrological loading. Whether the patterns in figure 2 are a physically meaningful expression of mass transports is a different question.

Spectral domain. Various signal and error spectra are represented in figure 3. The dashed TUM-1S signal spectrum is overlain over Kaula's curve. The TUM-1S error curve represents the error spectrum of the static field, derived from the full year of data. Every monthly solution will be worse by a factor $\sqrt{12}$, indicated by the dotted line.

The thin lines represent the signal spectra of the monthly solutions after subtraction of the mean. In principle, whenever a signal curve is above the error curve, it implies that this signal is recovered by CHAMP. One should be suspicious for two reasons,

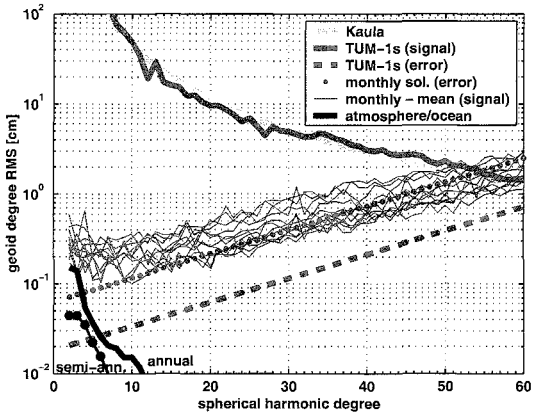


Fig. 3. Monthly CHAMP signal and error spectra compared to TUM1S and to seasonal gravity signal from atmosphere and ocean.

though. The monthly error spectrum is within or very close to the ensemble of monthly signal spectra. Moreover, the signal spectra are more or less parallel to the error spectrum, whereas one would expect a drop in power with increasing spherical harmonic degree. This leaves the meaningfulness of the patterns in figures 1 and 2 in doubt.

Figure 3 also contains two spectra with the annual and semi-annual signal power of combined oceanographically and meteorologically induced gravity variation. This combined model is derived from ECMWF pressure data and from MIT ocean circulation models over the time frame 01.1992–12.2000, cf. (Gruber & Peters, 2003). The semi-annual signal cannot be picked up in the monthly solutions. Only the annual signal curve rises above the monthly CHAMP error curve at the very lowest degrees.

A further analysis of time-series of individual residual spherical harmonic coefficients $\delta\bar{C}_{lm}(t)$ confirms the previous suspicion that the monthly geoid solutions are physically not meaningful. Figure 4 is built up of time-series plots of all $\delta\bar{C}_{lm}(t)$ coefficients up to $L = 5$ (multiplied by 10^{10}). Each black dot is a monthly (residual) coefficient. Also error bars are plotted. The monthly solutions are compared to grey curves, that represent annual and semi-annual variations derived from two years of SLR data from Cheng et al. (2003). The data set contains laser ranging to 6 laser satellites (Lageos-1, Lageos-2, Starlette, Ajisai, Stella and BEC) from January 2000 to December 2001.

Note that the time spans of the CHAMP and SLR solutions do not overlap. Nevertheless one would expect the annual and semi-annual signals to be consistent over a time frame longer than the one spanned by the SLR analysis. This expectation is confirmed, indeed, by the time series of residual coefficients in figure 5. This figure extends the SLR derived annual and semi-annual signals over the time frame of the aforementioned combined oceanographic-meteorological data set. For the selected coefficients, the correlation is convincingly high. For other coefficients—not shown here—the correlation can be lower.

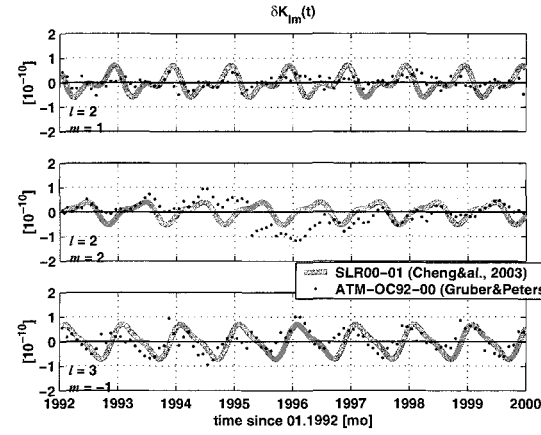


Fig. 5. Selected coefficients—(l, m)=(2,1), (2,2) and (3,-1)—of a combined oceanographic-meteorological model (01.1992–12.2000) compared to an annual and semi-annual SLR solution (01.2000–12.2001).

Thus, the SLR solution seems to reflect physically meaningful gravity field variations. Consequently, the present CHAMP results lack meaning. It can be concluded that the accuracy of the monthly CHAMP solutions is insufficient to recover low degree time variations. The main contribution to the error budget of the gravity solutions are the errors in the orbit data themselves. Apparently, even the kinematic orbits are not sufficient for these purposes.

Another main driver of the error budget is temporal and spatial aliasing. As mentioned before, time variations with periods shorter than a month will cause temporal alias into the monthly solutions. But also spatial aliasing, due to changing sampling configuration—read: changing ground tracks—will play an important role.

Orbit characteristics. CHAMP's orbit is slowly decaying. Apart from orbit manoeuvres on June 9–10 and December 9–10, 2002, the satellite gradually comes down, see figure 6. As a consequence the ground track pattern changes continuously. Figure 7 shows a dense coverage (two weeks in June

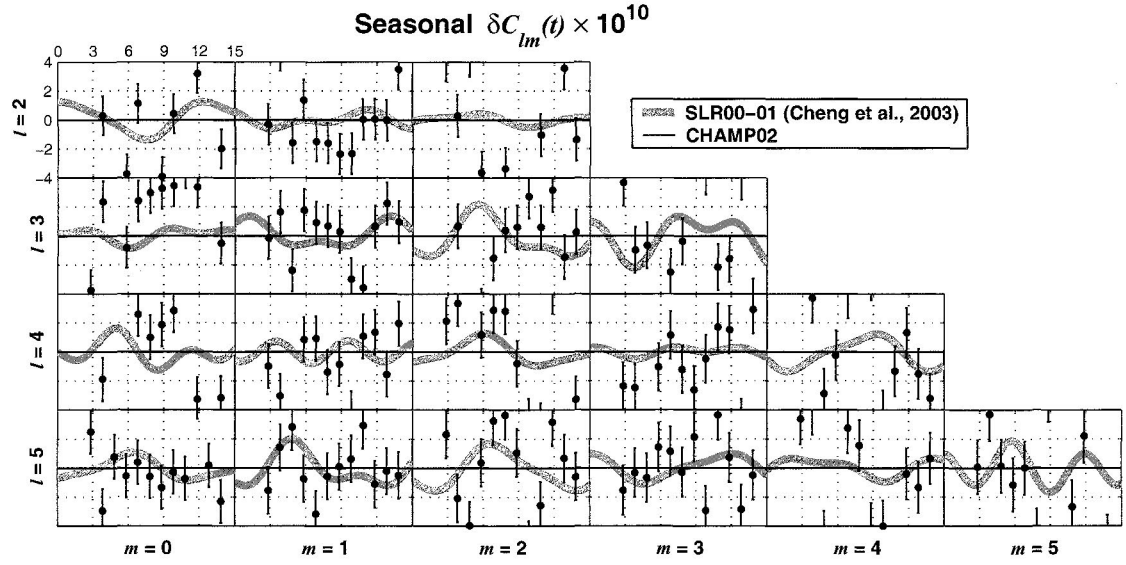


Fig. 4. Low degree spherical harmonic coefficients as time series (03.2002–02.2003) compared to annual and semi-annual SLR solution (01.2000–12.2001).

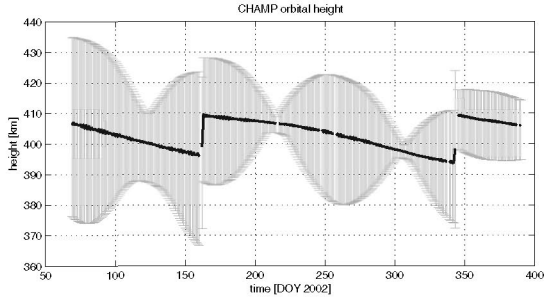


Fig. 6. CHAMP height profile in terms of daily mean/min/max. Note the orbit manoeuvres on June 9–10 and December 9–10.

2002) and a repeat-mode ground-track, that existed briefly before the first orbit manoeuvre. The satellite was effectively performing 31 revolutions in 2 days. The two orbit manoeuvres cause the satellite to go through this commensurability again twice.

In repeat mode, certain so-called lumped coefficients can be determined with high accuracy. In general, however, a repeat mode means low spatial resolution and a correspondingly low performance in overall gravity recovery. Figure 8 displays several repeat modes through which CHAMP can and will go. Later in its lifetime, at approximately 255 km height, the satellite will exactly perform 16 revolutions per day, leading to a very sparse spatial sampling.

4 Conclusions

One year of monthly CHAMP-only solutions from kinematic orbits has been derived using the energy balance approach. These orbits guarantee independence from a priori gravity field information. They are probably the best starting point for any CHAMP-only attempt to derive time-variable gravity.

Nevertheless, our attempt has not been successful so far. Although their magnitude is right, the monthly results do not seem to reflect physically meaningful gravity variations. Comparison to SLR-derived annual and semi-annual gravity changes and to gravity changes implied by a combined oceanographic-meteorological model has corroborated the above conclusion. Although the time frames for these three different data sets were distinct, the SLR solution was shown to be consistent with many, though not all, spherical harmonic coefficients of the oceanographic-meteorological model.

One reason for the lack of CHAMP sensitivity to time-variable gravity is the error level of the orbit data themselves. Most geophysical signal spectra are below the CHAMP error level. Also, the ground-track variability due to the orbit decay contributes greatly to the lack of sensitivity. Certain months show good spatial coverage. During other weeks or months, the satellite can be in repeat mode, leading to a loss of spatial coverage. One basically tries to solve a time-variable field from a time-variable sampling

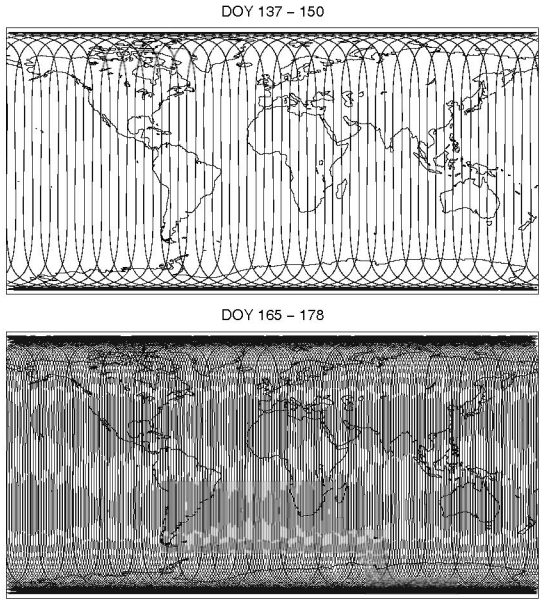


Fig. 7. Two-weekly ground track patterns: second half of May (top) and second half of June (bottom), 2002.

configuration. This spatial-temporal sampling problem comes on top of the temporal aliasing of short period effects into the monthly solutions.

The latter problem is discussed in the context of GRACE as well, e.g. (Velicogna et al., 2001). It is expected that ground-track variability will impede the consistency of the monthly GRACE solutions greatly.

References

Cheng, M. K., Gunter, B., Ries, J., Chambers, D. P., & Tapley, B. D. (2003). Temporal variations in the earth's gravity field from SLR and CHAMP GPS data. In I. N. Tziavos (Ed.), *Gravity and Geoid 2002* (pp. 424–431). 3rd meeting of the IGGC, Thessaloniki, Greece.

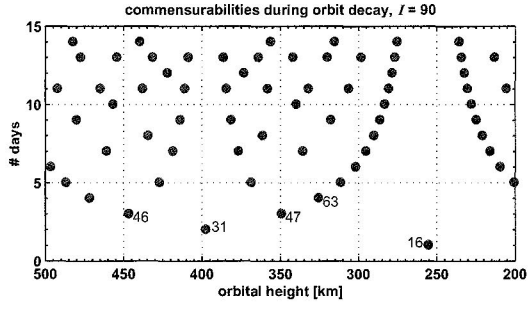


Fig. 8. Commensurabilities (repeat orbits) as a function of orbit height.

Gerlach, C., Földvary, L., Švehla, D., Gruber, T., Frommknecht, B., Oberndorfer, H., Peters, T., Rothacher, M., Rummel, R., Sneeuw, N., Steigenberger, P., & Wermuth, M. (2003a). A CHAMP only gravity field model from kinematic orbits using the energy balance approach. Poster presented at EGS-AGU-EUG Joint Assembly, 06–11 April 2003, Nice, France.

Gerlach, C., Földvary, L., Švehla, D., Gruber, T., Wermuth, M., Frommknecht, B., Peters, T., Rothacher, M., Rummel, R., Sneeuw, N., & Steigenberger, P. (2003b). A CHAMP only gravity field model from kinematic orbits using the energy integral. *Geoph. Res. Letters*. subm.

Gerlach, C., Sneeuw, N., Visser, P., & Švehla, D. (2003c). Champ gravity field recovery using the energy balance approach. *Adv. Space Res.*, 1, 73–80.

Gruber, T. & Peters, T. (2003). Identification of mass variations from a series of global gravity field models, a simulation study. Poster presented at IUGG 2003, Sapporo, Japan.

Han, S.-C., Jekeli, C., & Shum, C. K. (2002). Efficient gravity field recovery using in situ disturbing potential observables from CHAMP. *Geophys. Res. Letters*, 29(16). DOI 10.1029/2002GL015180.

Reigber, C., Schwintzer, P., Neumayer, K.-H., Barthelmes, F., König, R., Förste, C., Balmino, G., Biancale, R., Lemoine, J.-M., Loyer, S., Bruinsma, S., Perosanz, F., & Fayard, T. (2003). The CHAMP-only earth gravity field model EIGEN-2. *Adv. Space Res.* accepted.

Sneeuw, N. (2001). Satellite geodesy on the torus: Block-diagonality from a semi-analytical approach. In M. G. Sideris (Ed.), *Gravity, Geoid and Geodynamics*, volume 123 of *IAG Symposia* (pp. 137–142).: IAG Springer-Verlag. GGG2000, Banff, Canada.

Sneeuw, N., Gerlach, C., Švehla, D., & Gruber, C. (2003). A first attempt at time-variable gravity recovery from CHAMP using the energy balance approach. In I. N. Tziavos (Ed.), *Gravity and Geoid 2002* (pp. 237–242). 3rd meeting of the IGGC, Thessaloniki, Greece.

Švehla, D. & Rothacher, M. (2003a). Kinematic and reduced-dynamic precise orbit determination of CHAMP satellite over one year using zero-differences. Poster presented at EGS-AGU-EUG Joint Assembly, 06–11 April 2003, Nice, France.

Švehla, D. & Rothacher, M. (2003b). Kinematic and reduced-dynamic precise orbit determination of low earth orbiters. *Adv. Space Res.*, 1, 47–56.

Velicogna, I., Wahr, J., & Van den Dool, H. (2001). Can surface pressure be used to remove atmospheric contributions from GRACE data with sufficient accuracy to recover hydrological signals? *J. Geophys. Res.*, 106(B8), 16 415–16 434.

Visser, P. N. A. M., Sneeuw, N., & Gerlach, C. (2003). Energy integral method for gravity field determination from satellite orbit coordinates. *J. Geodesy*, 77, 207–216. DOI 10.1007/s00190-003-0315-8.

Topographic-Isostatic Reductions in Satellite Gravity Gradiometry Based on a Generalized Condensation Model

B. Heck, F. Wild

Geodetic Institute, University of Karlsruhe, Englerstr 7, D-76128 Karlsruhe, Germany
e-mail: heck@gik.uni-karlsruhe.de; Fax: +49-721-608-6808

Abstract. In satellite gravity gradiometry, the gravitational signals from the Earth's topography and its isostatic compensation still exist in the gravity gradients observed along the satellite orbit. Due to the high-frequency behaviour of the combined topographic-isostatic effect, downward continuation of the gravitational signal from satellite height to sea level is rather difficult, requiring some mathematical method of regularization. On the other hand, the complete calculation of topographic-isostatic effects according to, say, the Airy-Heiskanen isostatic model is too laborious for the practical evaluation of gravity gradiometry data. In this paper, another approach is proposed, which is based on a generalized condensation model corresponding to Helmert's first condensation model; here the condensed masses are assumed to be situated on a surface at a constant depth D below the geoid. The respective formulae representing the effects of the topographic and condensation masses on the vertical gravity gradient (V_{rr}) at satellite level are derived. A simulation based on the JGP95E rock-equivalent terrain model proves that the order of magnitude of both topographic and condensation effects is about 10 E.U. The magnitude of the combined topographic-condensation effect is much smaller, amounting to 0.6 E.U. and 0.06 E.U. for Helmert's first and second condensation model, respectively.

Keywords. Satellite gravity gradiometry, topographic reductions, Helmert's condensation models

1 Introduction

The Earth's topographic masses and its isostatic balance masses produce gravitational signals in external space, which will also be visible in satellite gravity gradiometry (SGG) observations, e.g. in the planned GOCE mission. Similar to terrestrial applications, it should be advantageous to reduce the effects of the topographic-isostatic masses in satellite gradiometry by some kind of remove-restore technique in order to produce a smooth field which can be downward continued more easily. The conventional procedure applied in terrestrial gravity field modelling relies on the computation of the gravitational effects of rectangular prisms – based on a detailed gridding of the topographical-isostatic masses (Tsoulis 1999; Kuhn 2003). Unfortunately the numerical effort is extremely high in this case.

In this paper, it is proposed to reduce these calculations to the evaluation of two-dimensional surface integrals. Assuming a constant topographic density, the respective volume integral can be analytically integrated over the height parameter, providing a two-dimensional spherical integral. On the other hand, taking Helmert's idea of condensation (see e.g. Heck 2003), a simple isostatic model can be constructed which behaves similarly to the Airy-Heiskanen isostatic model. The respective formulae for the evaluation of the second radial derivative of the topographic and isostatic potential constituents are derived in detail, and the numerical properties are outlined with the aid of a simulation.

2 Topographic Reductions of SGG Data

The gravitational potential of the topographical masses between sea level, approximated by the "geoidal sphere" Σ_g (radius R), and the Earth's topographical surface S is described by Newton's integral. Assuming a constant density ρ of the topographical masses and decomposing the triple integral into a two-dimensional spherical integral over σ and a one-dimensional integral over the radial parameter ξ results in

$$V_t(Q) = G\cdot\rho\cdot\iint_{\sigma}\left[\int_{R}^{r'}\frac{\xi^2}{\ell}d\xi\right]\cdot d\sigma \qquad (1)$$

$$\ell := \sqrt{r^2+\xi^2-2r\xi\cdot\cos\psi} \qquad (2)$$

where r is the radius of the computation point Q, ψ the angle between the geocentric radius of Q and the variable integration point P′ on S (see Figure 1), r′ the geocentric radius of P′, and G is the gravitational constant.

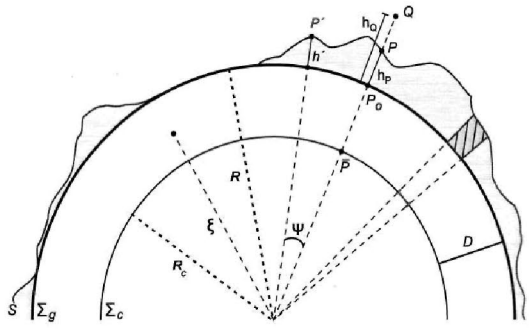


Figure 1. Geometry of the topography and the condensation model in spherical approximation

In Figure 1, $r = R + h_Q$ is the geocentric radius of the computation point Q, $r' = R + h'$ is the geocentric radius of the variable integration point P' $\in$ S, and $r_P = R + h_P$ the radius of the point P below Q at the intersection with the topographical surface S.

In Appendix A, the following expression for the second vertical (radial) derivative of the topographical potential V is proved:

$$\frac{\partial^2 V_t(Q)}{\partial r^2} = \frac{2}{r}\cdot\frac{\partial V_t(Q)}{\partial r} - \frac{2}{r^2}\cdot V_t(Q) + \frac{G\rho}{2r^2}\iint_{\sigma}\left[\frac{3\xi^3}{\ell}+\frac{\xi^3\left(r^2-\xi^2\right)}{\ell^3}\right]_{\xi=R}^{\xi=r'}\cdot d\sigma \qquad (3)$$

Formulae for $V_t(Q)$ and its first derivative have been derived by Heck (2003), see also Martinec (1998) and Novàk (2000), resulting in the formula

$$\frac{\partial^2 V_t(Q)}{\partial r^2} = \frac{G\rho}{r^2}\iint_{\sigma}\left[\begin{array}{l} -\frac{r'^3}{2\bar{\ell}'}+\frac{R^3}{2\tilde{\ell}}+r'\bar{\ell}'-R\tilde{\ell} \\ +\frac{r'^3\cdot\left(r^2-r'^2\right)}{2\bar{\ell}'^3}-\frac{R^3\cdot\left(r^2-R^2\right)}{2\tilde{\ell}^3} \\ +3r\cdot\cos\psi\cdot\left(\bar{\ell}'-\tilde{\ell}\right) \\ +r^2\cdot\left(3\cos^2\psi-1\right)\cdot\ell n\left|\frac{\bar{\ell}'+r'-r\cdot\cos\psi}{\tilde{\ell}+R-r\cdot\cos\psi}\right| \end{array}\right]\cdot d\sigma$$

where (4)

$$\bar{\ell}' = \sqrt{r^2+r'^2-2rr'\cdot\cos\psi}$$

$$\tilde{\ell} = \sqrt{r^2+R^2-2rR\cdot\cos\psi}$$

$$\cos\psi = \sin\phi\sin\phi'+\cos\phi\cos\phi'\cos(\lambda-\lambda').$$

Instead of integrating over the interval [R, r'] in one step, the interval can be decomposed into two parts [R, r_P] and [r_P, r'], where r_P is the radius of the point P below Q at the intersection with the topographical surface. An elementary evaluation of the respective integrals yields the alternative formula

$$\frac{\partial^2 V_t(Q)}{\partial r^2} = \frac{8\pi G\rho}{3r^3}\cdot\left(r_P^3-R^3\right)+ \frac{G\rho}{r^2}\iint_{\sigma}\left[\begin{array}{l} -\frac{r'^3}{2\bar{\ell}'}+\frac{r_P^3}{2\bar{\ell}}+r'\bar{\ell}'-r_P\bar{\ell}+\frac{r'^3\cdot\left(r^2-r'^2\right)}{2\bar{\ell}'^3} \\ -\frac{r_P^3\cdot\left(r^2-r_P^2\right)}{2\bar{\ell}^3}+3r\cdot\cos\psi\cdot\left(\bar{\ell}'-\bar{\ell}\right) \\ +r^2\cdot\left(3\cos^2\psi-1\right)\cdot\ell n\left|\frac{\bar{\ell}'+r'-r\cdot\cos\psi}{\bar{\ell}+r_P-r\cdot\cos\psi}\right| \end{array}\right]\cdot d\sigma$$

where (5)

$$\bar{\ell}' = \sqrt{r^2+r'^2-2rr'\cdot\cos\psi}$$

$$\bar{\ell} = \sqrt{r^2+r_P^2-2rr_P\cdot\cos\psi}\,.$$

The out-of-integral term in Eq. (5) can be interpreted as the effect of a spherical shell of constant density ρ and thickness $r_P - R$ on the vertical gravity gradient at Q, while the integral term corresponds to the influence of the variable terrain residual to this shell.

From an analytical point of view, the computation formulae Eqs. (4) and (5) are completely equivalent. However, the numerical properties of these formulae are quite different. Test calculations have shown that the major part of the

total topographical effect is provided by the out-of-integral term in Eq. (5), while the contribution of the integral term mostly is rather small. Thus, the integration over the spherical distance ψ may be truncated at some radius ψ_c, neglecting the far zone contribution. A straightforward application of Eq. (4) is not possible for the calculation of the contribution of the topographical column situated directly below the computation point Q, since the integrand contains terms of type 0/0 for $\psi \to 0$. Such problems do not occur in the evaluation based on Eq. (5), which therefore should be preferred in the numerical calculation of the topographical effects.

3 Generalized Condensation Model

Helmert's condensation models have been well-known in physical geodesy for more than 120 years, the first publication dating back to Helmert (1884). In his first method of condensation, Helmert proposed a redistribution of the topographical masses on an internal surface, 'parallel' to the geoid, located at a depth of 21 km below the geoid. In contrast, in the second method, the topographical masses are condensed on the geoid itself. The redistribution of masses is performed according to a local condensation procedure, i.e. a compression of the topographical column on the condensation surface. Helmert's second method of condensation has often been applied in recent years in practical geoid determination (Vaníček et al. 1999; Nahavandchi and Sjöberg 2001). A generalization of Helmert's first method in spherical approximation has been derived by Heck (2003), allowing the computation point Q to be situated outside the Earth's surface and fixing the (spherical) condensation layer at an arbitrary constant depth D below the 'geoidal sphere' Σ_g (see Figure 1). The potential of the masses condensed on the spherical surface Σ_c (radius R_c) running at the depth D below Σ_g can be expressed by the single layer potential

$$V_c(Q) = G \cdot \iint_\sigma \frac{\kappa'}{\ell_c} \cdot R_c^2 \cdot d\sigma ,$$
$$\ell_c = \sqrt{r^2 + R_c^2 - 2rR_c \cdot \cos\psi} . \qquad (6)$$

In order to fulfill the local mass conservation postulate, the surface density has to be chosen in the following way (Martinec 1998; Heck 2003)

$$\kappa' = \rho \cdot \frac{r'^3 - R^3}{3R_c^2} . \qquad (7)$$

Differentiating Eq. (6) twice with respect to r yields the second radial derivative

$$\frac{\partial^2 V_c(Q)}{\partial r^2} = \frac{G\rho}{r^2} \iint_\sigma \left[\frac{r'^3 - R^3}{4} \cdot \left(\frac{\left(r^2 - R_c^2\right)^2}{\ell_c^5} - \frac{4r^2}{3\ell_c^3} + 2\frac{r^2 - R_c^2}{\ell_c^3} + \frac{1}{\ell_c} \right) \right] \cdot d\sigma . \qquad (8)$$

An alternative representation can be achieved by adding and subtracting the constant density $\bar{\kappa} = \kappa(\bar{P})$, where $\bar{P} \in \Sigma_c$ is situated at the intersection of the radius vector of Q (see Figure 1):

$$\frac{\partial^2 V_c(Q)}{\partial r^2} = \frac{8\pi G\rho}{3r^3} \cdot \left(r_P^3 - R^3\right) + \qquad (9)$$
$$+ \frac{G\rho}{r^2} \iint_\sigma \left[\frac{r'^3 - r_P^3}{4} \cdot \left(\frac{\left(r^2 - R_c^2\right)^2}{\ell_c^5} - \frac{4r^2}{3\ell_c^3} + 2\frac{r^2 - R_c^2}{\ell_c^3} + \frac{1}{\ell_c} \right) \right] \cdot d\sigma .$$

Obviously, the out-of-integral term in Eq. (9) can be interpreted as the effect of a spherical layer with constant surface density $\bar{\kappa}$. Under the mass conservation postulate, it equals the corresponding term in Eq. (5). Thus, combining the effects of the topographical and condensation masses, the out-of-integral term cancels out in the difference between Eqs. (5) and (9), i.e.

$$\left[\frac{\partial^2 V_t(Q)}{\partial r^2} - \frac{\partial^2 V_c(Q)}{\partial r^2} \right] =$$

$$\frac{G\rho}{r^2} \iint_\sigma \left[\begin{array}{l} -\frac{r'^3}{2\bar{\ell}'} + \frac{r_P^3}{2\bar{\ell}} + r'\bar{\ell}' - r_P\bar{\ell} + \frac{r'^3 \cdot \left(r^2 - r'^2\right)}{2\bar{\ell}'^3} \\ -\frac{r_P^3 \cdot \left(r^2 - r_P^2\right)}{2\bar{\ell}^3} + 3r \cdot \cos\psi \cdot \left(\bar{\ell}' - \bar{\ell}\right) \\ + r^2 \cdot \left(3\cos^2\psi - 1\right) \cdot \ell n \left| \frac{\bar{\ell}' + r' - r \cdot \cos\psi}{\bar{\ell} + r_P - r \cdot \cos\psi} \right| \\ -\frac{r'^3 - r_P^3}{4} \cdot \left(\frac{\left(r^2 - R_c^2\right)^2}{\ell_c^5} - \frac{4r^2}{3\ell_c^3} + 2\frac{r^2 - R_c^2}{\ell_c^3} + \frac{1}{\ell_c} \right) \end{array} \right] \cdot d\sigma \qquad (10)$$

It should be noted that – after a suitable choice of the depth D of the condensation level – the results of the extended Helmert condensation model are

practically equivalent to the Airy-Heiskanen model; a detailed proof will be given in a forthcoming paper. The advantage of Helmert's model relies on the fact that only two-dimensional instead of three-dimensional integrals have to be evaluated.

4 Results

In order to investigate the properties of the derived procedure, a simulation based on the JGP95E digital elevation model (Lemoine et al. 1998) has been performed, involving a 1° x 1° grid of rock-equivalent topography. The radii of the mean Earth sphere (R), the concentric spheres of condensation (R_c) and the satellite (r) have been chosen in the following way:

$$R = 6371\,\text{km}$$

$$R_c = \begin{cases} R - 21\,\text{km} = 6350\,\text{km} & \text{(Helmert I)} \\ R - 0\,\text{km} = 6371\,\text{km} & \text{(Helmert II)} \end{cases}$$

$$r = R + 260\,\text{km} = 6631\,\text{km} \quad \text{(GOCE-like orbit)}.$$

In the numerical evaluation, the integration over any spherical 1° x 1° cell $\Delta\sigma_i$ has been replaced by summation

$$\iint_{\Delta\sigma_i} f(r, r', \psi) \cdot d\sigma \approx f(r, r'_m, \psi_m) \cdot \Delta\sigma_i ,$$

where r'_m, ψ_m refer to the centre of the cell $\Delta\sigma_i$.

In a first test, the equivalence of the numerical results based on the pair of equivations (4), (8) versus the set (5), (9) has been proved. In the sequel, Eqs. (5), (9) and (10) have been applied in the numerical calculations for the reasons discussed at the end of section 2. Furthermore, both Helmert I and II models have been compared.

The second radial derivative $\partial^2 V_t / \partial r^2$ of the topographic potential at the satellite height is illustrated in Figure 2. Maxima are visible in high-mountain regions, while minima occur near the deep-sea troughs. The total range of the signal is at the 10 E.U. level. At first sight, the results for the second radial derivative $\partial^2 V_c / \partial r^2$ of the condensation potential, corresponding to the Helmert I and II models, are practically the same and cannot be distinguished from Figure 2; the corresponding figures are not shown here.

The combined topographic-condensation effects for Helmert's first and second models are illustrated in Figures 3 and 4. It is obvious that both constituents partly cancel each other, such that the complete topographic-isostatic reduction of gravity gradients is one order of magnitude less in comparison to the single constituents in the case of the Helmert I model.

This cancellation property is even more pronounced for the Helmert II model (see Figure 4). The total topographic-isostatic reduction term is more than two orders of magnitude smaller than the single constituents.

Considering the fact that the isostatic masses compensating the Earth's topography are situated at a depth of about 20-30 km below sea level it is evident that Helmert's first model is a much better approximation to geophysical reality (Heck 2002). Thus, it becomes obvious that the topographic-isostatic reduction of gravity gradients is strongly underestimated when the reduction is calculated on the basis of Helmert's second model. As a consequence, the residual gravity gradient field – after having applied topographic-condensation reduction – will be much rougher when Helmert's second model is applied in the remove-restore process, and downward continuation is much more problematic.

5 Conclusions

It has become evident that topographic and isostatic effects are significant in satellite gradiometry, showing an order of magnitude of 10 E.U. in the second vertical derivatives of the potentials V_t and V_c at a satellite altitude of 260 km (approximately GOCE orbit). Topographic and isostatic effects corresponding to Helmert's first and second condensation models cancel out at a large extent. The combined effect amounts to about 0.6 E.U. for Helmert's first model, and 0.06 E.U. for Helmert's second model. Since Helmert's second model produces too low topographic-isostatic reduction terms, the residual field after having applied topographic-condensation reduction will remain rather rough; thus downward continuation of this field will be problematic.

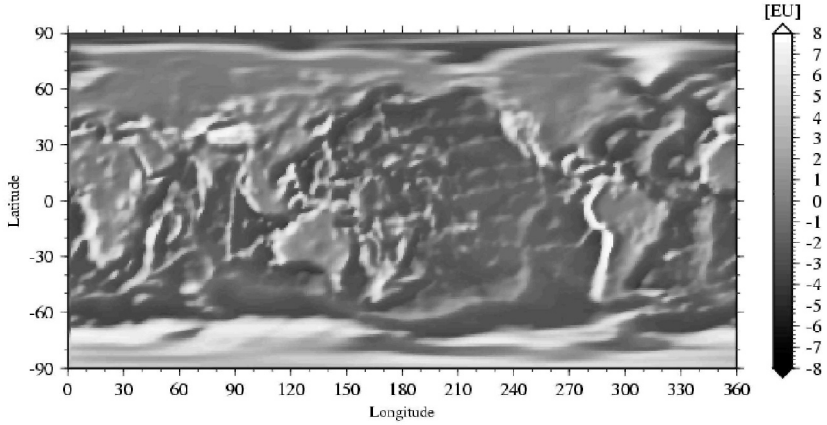


Figure 2. Second radial derivative $\partial^2 V_t / \partial r^2$ (nearly identical with $\partial^2 V_c / \partial r^2$)

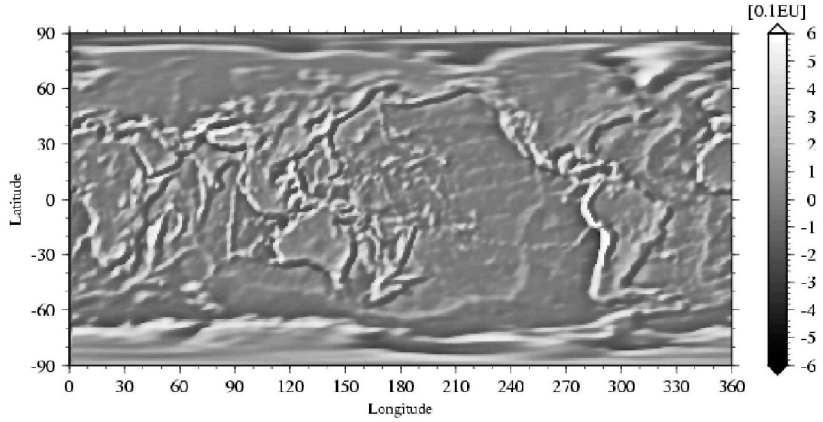


Figure 3. Topographic-condensation reduction $\left[\partial^2 V_t / \partial r^2 - \partial^2 V_c / \partial r^2\right]$ for the Helmert I model

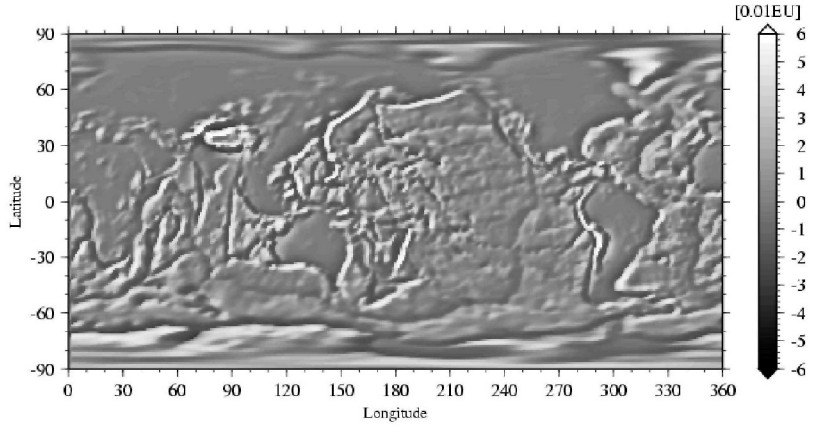


Figure 4. Topographic-condensation reduction $\left[\partial^2 V_t / \partial r^2 - \partial^2 V_c / \partial r^2\right]$ for the Helmert II model

Appendix A

In Heck (2003) the following identity has been derived:

$$\frac{\partial}{\partial r}\left(\frac{\xi^2}{\ell}\right)=\frac{2}{r}\cdot\frac{\xi^2}{\ell}-\frac{\partial}{\partial\xi}\left(\frac{\xi^3}{r\ell}\right),\ \ell=\sqrt{r^2+\xi^2-2r\xi\cdot\cos\psi}\,. \quad \text{(A1)}$$

A second differentiation with respect to r and elementary transformation of the right hand side yields

$$\frac{\partial^2}{\partial r^2}\left(\frac{\xi^2}{\ell}\right)=\frac{2}{r}\cdot\frac{\partial}{\partial r}\left(\frac{\xi^2}{\ell}\right)-\frac{2}{r^2}\left(\frac{\xi^2}{\ell}\right)+ +\frac{\partial}{\partial\xi}\left(\frac{3\xi^3}{2r^2\ell}+\frac{\xi^3\left(r^2-\xi^2\right)}{2r^2\ell^3}\right). \quad \text{(A2)}$$

This relationship can be used for deriving an analytical expression for the second radial derivative of the Newton integral (1):

$$\frac{\partial^2 V_t(Q)}{\partial r^2}=G\cdot\rho\cdot\iint_\sigma\left[\int_R^{r'}\frac{\partial^2}{\partial r^2}\left(\frac{\xi^2}{\ell}\right)d\xi\right]\cdot d\sigma$$

$$=G\cdot\rho\cdot\iint_\sigma\left[\int_R^{r'}\left(\frac{2}{r}\cdot\frac{\partial}{\partial r}\left(\frac{\xi^2}{\ell}\right)-\frac{2}{r^2}\left(\frac{\xi^2}{\ell}\right)+ +\frac{\partial}{\partial\xi}\left(\frac{3\xi^3}{2r^2\ell}+\frac{\xi^3\left(r^2-\xi^2\right)}{2r^2\ell^3}\right)\right)d\xi\right]\cdot d\sigma$$

$$=\frac{2}{r}\cdot\frac{\partial V_t(Q)}{\partial r}-\frac{2}{r^2}\cdot V_t(Q)+\frac{G\cdot\rho}{2\cdot r^2}\iint_\sigma\left[\frac{3\xi^3}{\ell}+\frac{\xi^3\cdot\left(r^2-\xi^2\right)}{\ell^3}\right]_{\xi=R}^{\xi=r'}\cdot d\sigma$$

Acknowledgements: The authors gratefully acknowledge the valuable comments by W. Featherstone und D. Tsoulis.

References

Heck B (2002) On the use and abuse of Helmert's second method of condensation. In: Ádám J, Schwarz KP (eds) Vistas for Geodesy in the new millennium. Springer IAG Symposia, Vol. 125

Heck B (2003) On Helmert's methods of condensation. Journal of Geodesy 77: 155-170

Helmert FR (1884) Die mathematischen und physikalischen Theorieen der Höheren Geodäsie. II. Teil: Die physikalischen Theorien, B.G. Teubner, Leipzig (reprinted 1962)

Kuhn M (2003) Geoid determination with density hypothesis from isostatic models and geological information. Journal of Geodesy 77: 50-65

Lemoine FG, Kenyon SC, Factor JK, Trimmer R, Pavlis NK, Chinn DS, Cox CM, Klosko SM, Luthcke SB, Torrence MH, Wang YM, Williamson RG, Pavlis EC, Rapp RH, Olson TR (1998) The development of the NASA GSFC and National Imaginary and Mapping Agency (NIMA) geopotential model EGM96. Rep. NASA/TP-1998-206861, National Aeronautics and Space Administration, Maryland

Martinec Z (1998) Boundary-value problems for gravimetric determination of a precise geoid. Lecture notes in Earth Sciences 73. Springer, Berlin Heidelberg New York

Nahavandchi H, Sjöberg LE (2001) Precise geoid determination over Sweden using the Stokes-Helmert method and improved topographic corrections. Journal of Geodesy 75: 74-88

Novàk P (2000) Evaluation of gravity data for the Stokes-Helmert solution to the geodetic boundary value problem. PhD Dissertation, Techn. Rep. of the Dept. of Geodesy and Geomatics Engineering, 207, University of New Brunswick, Fredericton

Tsoulis D (1999) Analytical and numerical methods in gravity field modelling of ideal and real masses. Deutsche Geodätische Kommission, C 510, Munich

Vanicek P, Huang J, Pagiatakis S, Véronneau M, Martinec Z, Featherstone W (1999) Determination of boundary values for the Stokes-Helmert problem. Journal of Geodesy 73: 180-192

Long-wavelength Discrepancies Between the Australian Gravity Field and EIGEN-2 CHAMP Data

W.E. Featherstone
Western Australian Centre for Geodesy, Curtin University of Technology, GPO Box U1987, Perth, WA 6845, Australia; fax +61-8-9266-2703; email W.Featherstone@curtin.edu.au

Abstract. Long-wavelength differences (in the ~1,100 km to ~6,500 km waveband) are evident between Australian free-air gravity anomalies and those implied by EIGEN-2 to degree 32, which is derived purely from CHAMP data. They reach ~11 mGal for the Australian land gravity anomalies and over ~15 mGal for the [unadjusted] ship-track gravity anomalies. These long-wavelength differences are used to quantify the effect on the gravimetric geoid model. Using an unmodified Stokes kernel and no limit on the cap radius (i.e., all terrestrial gravity in the region), the resulting effect on the geoid is between –0.72m and 0.13m, though the large negative value is attributed to a lack of gravity data to the north of Australia. Using a modified Stokes kernel and a one-degree cap radius as a high-pass filter, the resulting effect on the geoid is between –0.28m and 0.10m. Applying the latter and EIGEN-2 to degree 32 as corrections to AUSGeoid98 yields a small, but insignificant, improvement in the fit to nation-wide GPS-levelling data.

Keywords. Long-wavelength gravity anomaly errors, dedicated satellite gravimetry, CHAMP, geoid.

1 Introduction

Australian gravity anomalies are strongly suspected to contain long-wavelength errors due a combination of uncertainties in the gravity observations and their locations (cf. Heck 1990). This has adversely affected recent regional gravimetric geoid models (which are sensitive to long-wavelength terrestrial gravity anomaly errors; Vaníček and Featherstone 1998). As such, limited integration areas and/or modified integration kernels have been required to high-pass filter these errors (Forsberg and Featherstone 1998, Featherstone et al. 2001).

Pre-CHAMP satellite-only global geopotential models (GGMs) are limited in precision due to: power-decay of the Earth's gravitational field with altitude; inability to track complete satellite orbits from ground-based stations; imprecise modelling of atmospheric drag, non-gravitational and third-body gravitational perturbations; and incomplete sampling due to a limited number of orbital inclinations. Combined GGMs are limited in precision due to the above-mentioned issues, as well as the spatial coverage and quality of the terrestrial data used. Notably, they are subject to long-wavelength errors originating from the terrestrial gravity anomalies, and thus cannot be used to reliably isolate errors in (usually the same) terrestrial gravity data that have been used in the combined GGMs.

On the other hand, dedicated satellite gravimetry has begun a new era in regional gravity field determination (e.g., Rummel et al. 2002, Featherstone 2002, Tscherning et al. 2001). These missions will redress most of the above-mentioned limitations on previous satellite-only GGMs. However, the limitations on combined GGMs will remain unless the errors in terrestrial gravity anomalies can be properly removed (cf. Pavlis 1998, 2000). As such, modern dedicated satellite gravimetry has highlighted, and thus rejuvenated, the problems of long-wavelength errors in terrestrial gravity anomalies, and how these should be dealt with in regional geoid determination (e.g., Véronneau and Huang 2003).

This paper describes results of initial attempts to quantify long-wavelength errors in Australian gravity anomalies with a view to them being removed before a new Australian geoid is computed to supersede AUSGeoid98 (Featherstone et al. 2001). This involves computing differences between low-pass filtered terrestrial gravity anomalies and those implied by the degree-32 EIGEN-2 CHAMP-only GGM. It is acknowledged that improved CHAMP GGMs have become available very recently, as have the first GRACE GGMs. However, the studies reported here were conducted well before their release.

The long-wavelength differences, *cautiously* interpreted as terrestrial gravity anomaly errors, are then Stokes-integrated to yield the corresponding geoid error, which is then applied as part of a correction to AUSGeoid98 as an interim solution.

2. EIGEN-2 over Australia

EIGEN-2 (Reigber *et al.* 2003) is a GGM derived purely from CHAMP high-low satellite-to-satellite tracking and on-satellite accelerometry data. Further details of the CHAMP mission and derived GGMs are at http://op.gfz-potsdam.de/champ/.

EIGEN-2 only contains full signal power (i.e., degree variance) to degree ~40, accompanied by a corresponding increase in signal noise (i.e., error degree variance). This was ascertained from comparisons with EGM96 (Lemoine et al. 1998), acknowledging the restrictions on combined GGMs. EIGEN-2 was truncated at degree 32 for this study because this is the point beyond which its error degree variance exceeds that of EGM96. Over Australia (mean latitude ~27°S), this equates to wavelengths longer than ~1,113 km. The dimensions of the study area also set a maximum resolvable wavelength of ~6,500 km. Figure 1 shows degree-32 EIGEN-2 gravity anomalies over Australia with respect to GRS80 (Moritz 1980).

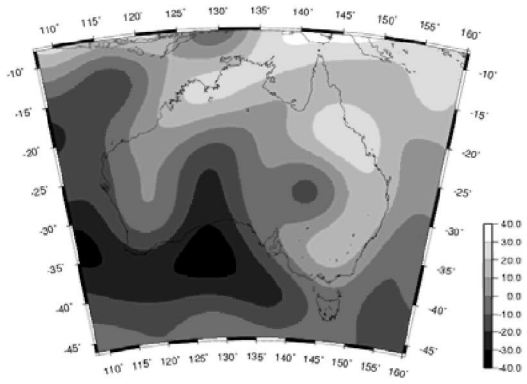


Fig 1. Long-wavelength gravity anomalies implied by EIGEN-2 to degree 32 (mGal; Lambert projection)

The internally estimated error in EIGEN-2 is <~0.5 mGal (Reigber et al. 2003). However, this is probably too optimistic: comparisons of EIGEN-2 with the UCPH series of CHAMP-derived GGMs (http://www.gfy.ku.dk/~cct/sagrada.htm) show differences of up to 6 mGal over Australia. Given that these GGMs were derived from largely the same data, this substantiates the comment made by Lambeck and Coleman (1983) "*...the various models are not as good as they are said to be. If they were, the differences between them should not be so great as they are...*". As such, extreme caution should be exercised if using these GGMs to claim errors in terrestrial gravity anomalies.

3. The Australian Gravity Data

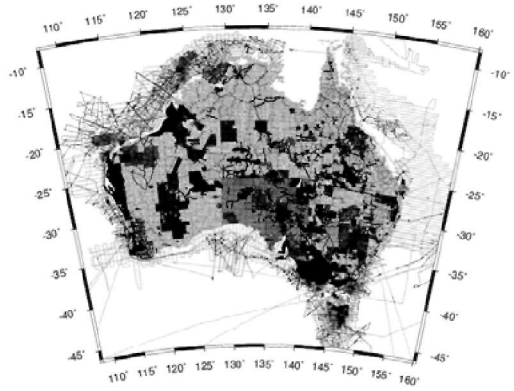


Fig 2. Spatial coverage of the 809,501 land and 51,141 [unadjusted] marine gravity observations in the 2002 data release from Geoscience Australia

3.1 Land gravity data

A total of 809,501 land gravity observations were taken from the 2002 release of Geoscience Australia's (GA's) gravity database (Murray 1997). The overall spatial resolution of the data is ~7-11 km, but this decreases in areas of resource prospectivity or scientific interest (Fig 2). The gravity observations refer to the IsoGal84 Australian gravity datum (Wellman *et al.* 1985), which is tied to the IGSN71 (Morelli *et al.* 1971). The tidal system is unknown. For the ~7-11 km Australian reconnaissance gravity (Fraser *et al.* 1976), the observation errors have been estimated as ~±0.3 mGal (Barlow 1977). More recent gravity observations with new gravimeters may be as precise as ~±0.01 mGal.

The elevations of the GA gravity observations are of a highly variable quality. For the reconnaissance gravity data, the error in the barometrically determined elevations has been estimated to be 4-6m (Barlow 1997), but may exceed 10m for heights scaled from topographic maps. Moreover, part of the reconnaissance gravity survey was conducted before the establishment of the Australian Height Datum (AHD; Roelse *et al.* 1971). Subsequent surveys have been tied to monumented stations of this reconnaissance survey and existing AHD benchmarks. More recently, GPS and [unspecified] geoid models have been used to coordinate regional gravity surveys. Therefore, much of the Australian gravity data is not rigorously tied to the AHD, which affects the accuracy and consistency of the gravity anomalies (e.g., Featherstone *et al.* 2001).

Taking the above factors together with those discussed by Heck (1990), Australian land gravity anomalies can reasonably be expected to contain

long-wavelength errors of a few (say, ~3-5) mGal. As stated in the Introduction, these appear to have degraded geoid computations over Australia.

3.2 Ship-track gravity data

Ship-borne gravity measurements are inherently less precise than land-based observations due to accelerations of the observing platform and inaccurate navigation, which affects the computation of normal gravity and Eötvös corrections, as well as incorrectly mapping the anomalies. Systematic errors occur along the ship-tracks due to a combination of gravimeter drift with infrequent crossovers, notably for oceanic cruises (Wessel and Watts 1998). In addition, ship-borne gravity data are of limited spatial resolution and concentrate in areas of resource prospectivity or scientific interest (Fig 2).

When marine gravity measurements are made progressively further away from ports, the accumulated effect of gravimeter drift cannot be completely removed by crossover adjustments. Instead, the errors are 'smeared out' among the ship-tracks, resulting in long-wavelength errors. As such, marine gravity anomalies can be expected to contain larger long-wavelength errors, especially towards the outer edges of the survey area and further away from ports.

For this study, a set of 51,441 marine gravity observations (Fig 2) along reasonably well-navigated ship tracks (though the accuracy of navigation varies among cruises) was used. However, they have not yet been crossover adjusted (Featherstone 2003). Unfortunately, this was not known at the time AUSGeoid98 was computed, apart from some grossly erroneous oceanic cruises to the southwest of Australia (Featherstone *et al.* 2001). Accordingly, AUSGeoid98 should be treated sceptically in these marine regions (cf. Fig 2).

Satellite altimeter-derived gravity anomalies are also available in the Australian region. Featherstone (2003) used these to verify errors in the ship track data, showing that they had not been crossover adjusted. However, satellite altimeter-derived gravity anomalies have not been used here because they are usually based on EGM96. Therefore, the long-wavelength differences between them and EIGEN-2 are largely the same as the degree-32 differences between EGM96 and EIGEN-2 (cf. Fig 5).

4. Method, Results and Discussion

Second-order free-air, atmospherically corrected terrestrial gravity anomalies were recomputed with respect to GRS80 (e.g., Featherstone and Dentith 1997). The degree-32 EIGEN-2 gravity anomalies were evaluated for the geocentric latitude and longitude of each terrestrial gravity anomaly, and then subtracted. Since the terrestrial gravity anomalies contain all frequencies, a 1,113 km low-pass cosine-arch-tapered filter was used to show only the long-wavelength differences commensurate with a degree-32 expansion. The filter was chosen from the GMT (Generic Mapping Tools) library, and gave similar results to other filters.

4.1 Land gravity anomaly differences

From Fig 3, long-wavelength (1,100 km–6,500 km) differences of up to ~11 mGal in magnitude exist in the gravity anomalies over the Australian landmass, with the majority being between −1 mGal and +7 mGal. Comparing Fig 3 with Fig 2 shows that there is generally a high level of correlation of these differences with the spatial coverage of the observations. However, these differences are large in relation to the expected few mGal errors in the Australian land gravity anomalies. Therefore, it is plausible that some of this difference is due to long-wavelength errors in EIGEN-2. This will probably only be verified when new GGMs are released based on additional CHAMP data, as well as the GRACE and GOCE satellite gravity missions.

Notwithstanding the errors in EIGEN-2, it is likely that the gravity anomaly errors on land have been caused by incorrect positioning (notably in remote regions that were surveyed pre-AHD) and ties among gravity surveys, as well as errors in the IsoGal84 base-station network that have propagated into regional surveys. Some of the large difference over the eastern goldfields in Western Australia (~28°S, ~121°E) could be caused by the use of GPS to coordinate the more recent spatially dense gravity surveys. As stated earlier, the geoid model used to transform the GPS ellipsoidal heights to the AHD is currently unspecified in the GA database. Other long-wavelength differences in Fig 3 are centred at: ~25°S, ~115°E, which could be due to errors in the GPS-controlled gravity surveys conducted here; ~25°S, ~134°E, which could be due to the in-fill surveys in South Australia; ~17°S, ~131°E and ~33°S, ~144°E, which could be due to the sparse reconnaissance surveys in these remote regions.

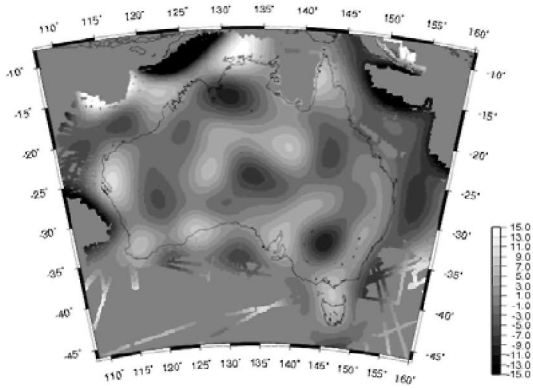


Fig 3. Low-pass cosine filtered (1,113 km) differences between Australian land and marine and degree-32 EIGEN-2 gravity anomalies (mGal)

4.2 Marine gravity anomaly differences

The long-wavelength differences in the [unadjusted] ship-track gravity anomalies are much more significant, exceeding 15 mGal in magnitude over some regions (Fig 3). These larger differences tend to occur beyond ~400 km offshore Australia, which are attributed to the propagation of gravimeter drift as the survey vessels move further away from the coast (discussed earlier).

To the north of Australia (notably ~12°S, ~125°E), however, these large differences are also due to the Gibbs phenomenon acting on the low-pass filter because the marine gravity anomalies tend to be larger towards Indonesia and offshore Western Australia. The larger difference in the Gulf of Carpentaria (~18°S, ~152°E) may be due to a combination of the Gibbs phenomenon with extrapolation of the gravity anomalies surrounding this marine data void. As expected, the Gibbs phenomenon is not as large in other areas where the gravity anomalies are smaller in magnitude. At present, however, the larger long-wavelength differences away from the coast cannot be reliably isolated between errors and the Gibbs phenomenon.

In the near-coastal regions (<~400 km), the long-wavelength differences are typically less than ~10 mGal. These are generally correlated with the edge of the continental shelf, except to the north of Australia. Finally, large long-wavelength differences are often apparent in the isolated oceanic cruises that are not observed in the familiar hashed pattern (Fig 2; also see Featherstone 2003). This is as expected because these cruises are more subject to unmitigated gravimeter drift (cf. Wessel and Watts 1998).

4.3 Effects on the gravimetric geoid model

The long-wavelength differences in Fig 3 (interpreted cautiously here as terrestrial gravity anomaly errors) were then used to quantify the effect on the regional gravimetric geoid model, first using the unmodified Stokes kernel and no limit on the cap radius (i.e., all land and marine gravity anomalies in the region). The unmodified Stokes kernel is used with no cap radius because, unlike some modified kernels with limited cap radii, it does not efficiently high-pass filter terrestrial gravity anomaly errors from the combined geoid solution (Vaníček and Featherstone 1998).

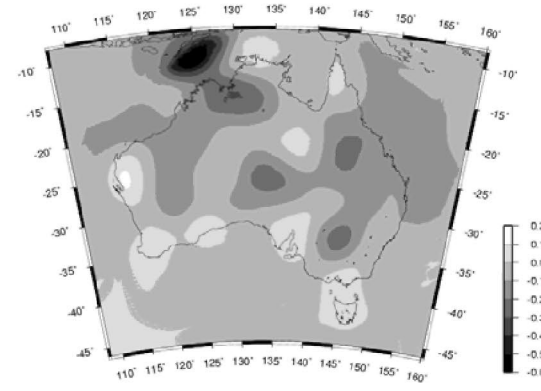


Fig 4. Geoid errors due to long-wavelength terrestrial gravity anomaly 'errors' (Fig 3), computed with the unmodified Stokes kernel over the whole data area (metres)

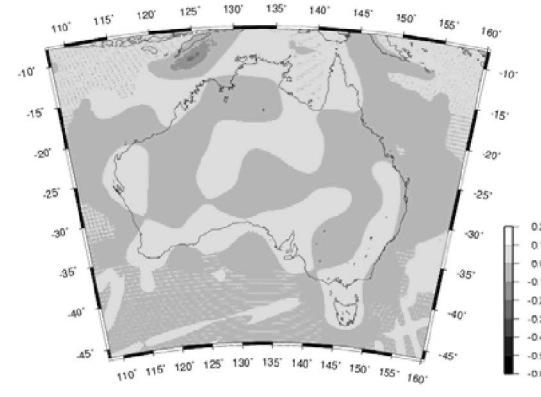


Fig 5. Geoid errors due to long-wavelength terrestrial gravity anomaly 'errors' (Fig 3), computed with the modified kernel over a one-degree spherical cap (metres)

The unmodified kernel represents a somewhat worst-case scenario of how long-wavelength terrestrial gravity anomaly errors may affect the Austra-

lian geoid (Fig 4). The large negative value (–0.72 m) just to the south of Indonesia (~12°S, ~125°E) in Fig 4 is attributed to the large negative marine gravity anomaly error (cf. Fig 3), though this is due to the Gibbs phenomenon in this region (see the earlier discussion). Over the Australian landmass, the long-wavelength geoid error implied via the unmodified kernel is typically <0.3 m (Fig 4).

A second geoid computation was performed using the Featherstone et al. (1998) deterministically modified kernel over a one-degree integration radius, which were used for AUSGeoid98 (Featherstone *et al.* 2001); see Fig 5. The aim of this was twofold: 1) to permit a self-consistent 'geoid correction' to be applied to AUSGeoid98; and 2) to empirically demonstrate the high-pass filtering properties of the modified kernel in relation to the unmodified kernel (cf. Vaníček and Featherstone 1998). Comparing Figs 4 and 5 clearly demonstrates the high-pass filtering properties of the modified kernel and limited integration radius (also see Table 1). For instance, the geoid feature to the south of Indonesia is now diminished to –0.28 m.

Geoid error	*Max*	*Min*	*Mean*	*STD*
Unmodified kernel, all data	0.132	-0.722	-0.069	±0.079
Modified kernel, one-degree cap	0.103	-0.276	-0.002	±0.018
EGM96 minus EIGEN-2 (n=32)	0.292	-0.464	-0.008	±0.110

Table 1. Statistics of the long-wavelength geoid errors using different kernels and integration areas, and geoid differences between degree-32 EGM96 and EIGEN-2 geoids (metres)

4.4 Correction to AUSGeoid98

AUSGeoid98 is based on EGM96, and is thus subject to long-wavelength errors in this GGM, as well as from the propagation of long-wavelength terrestrial gravity anomaly errors, though these will have been attenuated by the modified kernel and limited spherical cap. Therefore, a 'geoid correction' was attempted for AUSGeoid98 by:

1) Replacing the degree-32 EGM96 geoid with the degree-32 EIGEN-2 geoid (Fig 6 and Table 1)
2) Applying the 'geoid errors' given by the modified Stokes integration (Fig 5 and Table 1).

The resulting 'corrected' AUSGeoid98 was then compared with a nation-wide set of 1,013 GPS-levelling data (Table 2). Comparing Figs 5 and 6 (cf. Table 1) shows that the differences between EGM96 and EIGEN-2 are the more significant part of the 'geoid correction'. This is to be expected because most of the geoid's power is contained in the low frequencies, whereas Stokesian integration over a limited area cannot provide all low frequencies (Vaníček and Featherstone 1998).

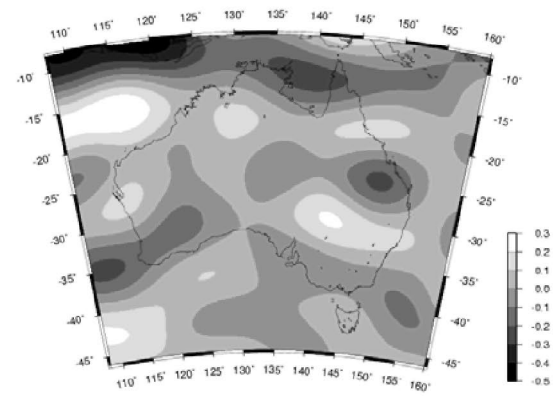


Fig 6. Differences between EGM96 and EIGEN-2 geoid heights to degree 32 (metres)

Geoid model	*Max*	*Min*	*Mean*	*STD*
AUSGeoid98	3.558	-2.572	-0.002	±0.314
'Corrected' AUSGeoid98	3.408	-2.637	-0.001	±0.308

Table 2. Statistics of the differences between gravimetric geoid models and 1,013 GPS-levelling data (metres)

Table 2 shows that the total 'geoid correction' slightly (but not statistically significantly, given the likely errors in the GPS-levelling data; say, ~5-10 cm) improves AUSGeoid98 in terms of the fit to GPS-levelling data. However, it is important to remember that the GPS data are of variable quality, and the levelling is tied to the AHD, which is strongly suspected to contain 1-2 m distortions.

5. Summary and Conclusions

This paper has used the degree-32 expansion of EIGEN-2, a GGM derived purely from CHAMP dedicated satellite gravimetry data, to map and quantify suspected long-wavelength errors (within the ~1,100 km to ~6,500 km waveband) in land and marine free-air gravity anomalies over Australia. This comparison shows differences of up to ~11 mGal magnitude in free-air gravity anomalies on land and of over ~15 mGal magnitude in [unadjusted] ship-track gravity anomalies. However, these differences cannot be interpreted solely as

terrestrial gravity anomaly errors because of uncertainties in EIGEN-2.

Nevertheless, these differences were then Stokes-integrated using 1) the unmodified kernel and no limit on the cap radius, and 2) a deterministically modified kernel and a one-degree spherical cap radius (i.e., the same as AUSGeoid98). The former approach shows a maximum effect on the geoid of 0.72m in magnitude, whereas the latter approach reduces it to 0.28m in magnitude. This empirically demonstrates the high-pass filtering properties of the modified kernel and limited spherical cap.

The long-wavelength geoid 'error' from the modified Stokes integration was combined with the difference between the degree-32 EGM96 and EIGEN-2 geoids, and then applied to AUSGeoid98 as a 'geoid correction' term. This only marginally improved the fit to a nation-wide set of 1013 GPS-levelling data, but this could equally be attributed to errors in the GPS-levelling data. Future studies will continue to attempt to identify and remove long-wavelength errors from the Australian terrestrial gravity anomalies with a view to improved regional geoid models. This will include data from refined CHAMP GGMs and new GGMs derived from the GRACE and GOCE dedicated satellite gravimetry missions.

Acknowledgments. *I would especially like to thank the many providers of data, who are either named or cited in this paper. This research is funded as part of Australian Research Council discovery grant DP0211827. Thanks also go to the two reviewers for their very constructive critiques of an earlier version of this paper.*

References

Barlow BC (1977) Data limitations on model complexity; 2-D gravity modelling with desktop calculators, Bull Aust Soc Expl Geophys 8: 139-143

Featherstone WE (2002) Expected contributions of dedicated satellite gravity field missions to regional geoid determination with some examples from Australia, J Geospatial Eng 4(1): 2-19

Featherstone WE (2003) Comparison of different satellite altimeter-derived gravity anomaly grids with ship-borne gravity data around Australia, in: Tziavos IN (ed) Gravity and Geoid 2002, Aristotle Univ of Thessaloniki, 326-331

Featherstone WE, Dentith MC (1997) A geodetic approach to gravity reduction for geophysics, Comp & Geosci 23: 1063-1070

Featherstone WE, Evans JD, Olliver JG (1998) A Meissl-modified Vaníček and Kleusberg kernel to reduce the truncation error in gravimetric geoid computations, J Geod 72: 154-160

Featherstone WE, Kirby JF, Kearsley AHW, Gilliland JR, Johnston GM, Steed J, Forsberg R, Sideris MG (2001) The AUSGeoid98 geoid model of Australia: data treatment, computations and comparisons with GPS-levelling data, J Geod 75: 313-330

Forsberg R, Featherstone WE (1998) Geoids and cap sizes, in: Forsberg R, Feissl M, Dietrich R (eds) Geodesy on the Move, Springer, Berlin, 194-200

Fraser AR, Moss FJ, Turpie A (1976) Reconnaissance gravity survey of Australia, Geophysics 41: 1337-1345

Heck B (1990) An evaluation of some systematic error sources affecting terrestrial gravity anomalies, Bull Géod 64: 88-108

Lambeck K, Coleman R (1983) The Earth's shape and gravity field: a report of progress from 1958 to 1982, Geophys J Royal astr Soc 74: 25-54

Lemoine FG and 14 other authors (1998) The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) geopotential model EGM96, NASA/TP-1998-206861, NASA, Washington

Morelli C, Gantar C, Honkaslo T, McConnell RK, Tanner TG, Szabo B, Uotila U, Whalen CT (1971) The International Gravity Standardisation Network 1971, Spec Pub 4, Bull Géod, Paris

Moritz H (1980) Geodetic reference system 1980, Bull Géod 54: 395-405

Murray AS (1997) The Australian national gravity database, AGSO J Aust Geol & Geophys 17: 145-155

Pavlis NK (1998) Observed inconsistencies between satellite-only and surface gravity-only geopotential models, in: Forsberg R, Feissl M, Dietrich R (eds) Geodesy on the Move, Springer, Berlin, 144-149

Pavlis NK (2000) On the modelling of long wavelength systematic errors in surface gravimetric data, in: Schwarz KP (ed) Geodesy Beyond 2000, Springer, Berlin, 131-136

Reigber C, and 12 other authors (2003) The CHAMP-only Earth gravity field model EIGEN-2, Advances in Space Research 31(8): 1883-1888

Roelse A, Granger HW, Graham JW (1971) The adjustment of the Australian levelling survey 1970–1971, Tech Rep 12, Div of National Mapping, Canberra, 81 pp.

Rummel R, Balmino G, Johannessen J, Visser P, Woodworth P (2002) Dedicated gravity field missions - principles and aims, J Geodynamics 33: 3-20

Tscherning CC, Arabelos D, G Strykoski G (2001) The 1 cm geoid after GOCE, in: Sideris MG (ed) Gravity Geoid and Geodynamics 2000, Springer, Berlin, 267-270

Vaníček P, Featherstone WE (1998) Performance of three types of Stokes's kernel in the combined solution for the geoid, J Geod 72: 684-697

Véronneau M, Huang J (2003) Correction of systematic errors in surface gravity anomalies using satellite gravity solutions, in Tziavos IN (ed) Gravity and Geoid 2002, Aristotle Univ of Thessaloniki, 217-222

Wellman P, Barlow BC, Murray AS (1985) Gravity base station network values, Rep 261, Australian Geological Survey Organisation, Canberra.

Wessel P, Watts AB (1988) On the accuracy of marine gravity measurements, J Geophys Res 94: 7685-7729

Comparison of new Geoid Models and EIGEN-2S in the North Atlantic Region

O. C. D. Omang
Dept. of Mathematical Sciences and Technology, NLH, N-1432 Ås, Norway; e-mail:ove.omang@nlh.no

R. Forsberg, G. Strykowski
Geodynamic Dept., KMS, Rentemestervej 8, DK-2400 Copenhagen NV, Denmark; e-mail: rf@kms.dk, gs@kms.dk

Abstract. A set of new gravity data, not previously available in the Nordic NKG data base, was obtained from NIMA . Based on existing gravity data and these additional data several different new geoids were estimated. The inclusion of the NIMA data set give local geoid changes of up to a few dm.

The new CHAMP based EIGEN-2S geopotential model was compared to the EGM96 model using a two step procedure. Results from the two steps show that the EIGEN-2S geopotential model give better fit to GPS and MSSH/MDT data than EGM96 up to degree 50, but applying the full EGM96 model still gives the best resulting geoid.

Keywords. EIGEN-2S – new Geoid models – North Atlantic region

1 Introduction

Large parts of the test area is covered with good quality gravity data at a reasonable density level. However, in the Greenland/Norwegian sea and south of Iceland there are almost no data available, see Fig. 1. Usually these areas have been filled with gravity data obtained from satellites such as ERS-2.

Recently, new gravity data covering the North Atlantic region was kindly provided by National Imagery and Mapping Agency (NIMA). As seen in Fig. 2 the new data set improves the gravity coverage in the North Atlantic region. Based on these data, a set of new geoids have been calculated to estimate the effect of adding the new gravity data to the existing gravity data set.

As a result of the dedicated gravity satellite mission CHAMP, new geopotential models have been derived. These models are expected to give a better representation of the long wavelength part of the Earth's gravity field than previous geopotential models. A two step test procedure was executed to compare the EIGEN-2S geopotential model to the EGM96 model. In the first step, a set of geoids, based on the geopotential models only, were created by varying the spherical harmonic degree, N_{max}. In the second step, a gravimetric geoid was produced by combining the *best* geopotential model and available gravity data. A set of these geoids were created by modifying Stokes function. The resulting geoids were compared to GPS networks in the Nordic countries and to a combination of mean sea surface heights (MSSH) and Mean Dynamic Topography (MDT) data over the ocean.

The test area covered the North Atlantic Sea region, including the Greenland Sea and the Norwegian Sea.

1.1 The *best* geoid

In this context the *best* geoid means the geoid which gives the minimum std. dev. to a selected dataset, i.e. MSSH/MDT or GPS data. In the case where the comparison is with MSSH/MDT data, the geoid with the best fit does not necessarily represents the correct geoid. This is due to the fact that the sea surface topography is affected by other factors than gravity. These factors may be sea currents, waves, tides, etc.

2 Global geopotential models

In this paper the following global geopotential models were used:

- The Earth Gravity Model (1996) (EGM96) (Lemoine *et al.*, 1996)
- The CHAMP based model EIGEN-2S (Reigber *et al.*, 2003)

3 Free-air anomalies

The free-air anomalies were obtained from four different sources: the Nordic Geodetic Commission (NKG) gravity data base, the Arctic Gravity Project (ArcGP), the Kort og Matrikelstyrelsen satellite altimetry gravity (KMS02) and the NIMA data set. Statistics of the free-air anomalies are listed in Table 1.

Table 1. Statistics of the selected data points; free-air anomalies in mGal

Name	No points	Min	Max	Mean	Std.dev
NKG	144265	-226.450	245.740	9.217	31.524
ArcGP	31689	-185.200	226.600	23.706	37.989
NIMA	28669	-59.400	174.500	26.913	26.056
KMS02	30195	-61.660	117.385	23.291	20.482

- NKG: Figure 1 illustrates the distribution of the selected gravity points, taken from the NKG gravity data base, within the geographical area 58°-83°N, 40°W-32°E. All available NKG data points within this area were merged into gridcells of size 0°.05 x 0°.1, reducing the total number from approximately 500 000 to 144 265 points. NKG gravity data over Greeland, Iceland and Svalbard were substituted with ArcGP data.
- ArcGP: Illustrated in Fig. 1 are the free-air anomalies selected from the Arctic Gravity Project. The original data have been regridded to 0°.05 x 0°.1. These data cover the land areas of Greenland, Iceland, Svalbard and the geographical region 82°N-83°N, 2°W-32°E.
- NIMA: Illustrated in Fig. 2 are the selected NIMA free-air anomalies (Kenyon, 2003). NIMA data was selected in areas with no NKG data available within 3 km.
- Kort og Matrikelstyrelsen (KMS)02: The KMS02 data (Andersen and Knudsen, 1998) was selected over ocean areas only, where no other gravity data was available within 20 km.

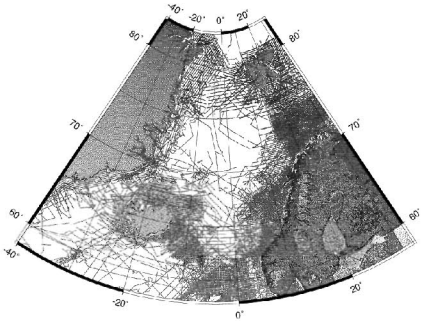


Fig. 1. Distribution of the NKG data. The ArcGP data covers the land areas of Greenland, Iceland, Svalbard and the geographical region 82°N-83°N, 2°W-32°E

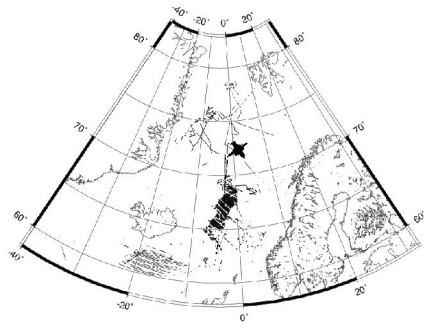


Fig. 2. Distribution of the selected NIMA free-air gravity data

4 GPS, MSSH and MDT data

The global geoid models and the local gravimetric geoids are compared both to GPS stations and the MSSH-MDT data.

The GPS network covers the whole Fennoscandia area including the Baltic. A total of 173 GPS stations are used in the comparison.

Generated MSSH data is available from several sources, e.g. KMS and Collecte Localisation Satellites (CLS), and give the mean sea surface heights within 82°S — 82°N. Our data set is from KMS (Andersen, 2003). Data along coastlines and especially in areas around the north eastern part of Greenland was removed from the selected dataset by choosing a cutoff error level of 0.04, see Fig. 4.

As Mean Dynamic Topography (MDT) the Ocean Circulation and Climate Advanced Modelling (OCCAM) model was used with the same cutoff level as MSSH. The combined result, MSSH *minus* MDT, in principle gave the geoid.

Statistics are listed in Table 2.

Table 2. Statistics of the GPS network and the MSSH, MDT and MSSH *minus* MDT (M-M) data set; values in m

Name	No points	Min	Max	Mean	Std.dev
GPS	173	14.510	46.730	27.544	9.284
MSSH	171450	15.650	66.360	46.922	11.894
MDT	171450	-0.991	0.158	-0.577	0.166
M-M	171450	15.492	66.905	47.499	11.944

5 Computational formulae

The gravimetric geoids are calculated using the remove-restore (R-R) method. First, the global geopotential

Table 3. Statistics of the different EGM96 geoids compared to GPS network and to MSSH-MDT; values in m

N_{max}	GPS Mean	GPS SD	MSSH-MDT Mean	MSSH-MDT SD
20	-0.904	2.852	0.083	2.105
40	-0.838	1.304	-0.047	1.291
50	-0.805	1.315	-0.008	1.052
60	-0.712	1.178	0.020	0.885
80	-0.858	0.997	0.063	0.773
120	-0.648	0.562	0.049	0.562
360	-0.566	0.240	0.051	0.253

Table 4. Statistics of the different EIGEN-2S geoids compared to GPS network and to MSSH-MDT; values in m

N_{max}	GPS Mean	GPS SD	MSSH-MDT Mean	MSSH-MDT SD
20	-0.863	2.864	0.064	2.094
40	-0.859	1.281	-0.059	1.243
50	-0.828	1.253	-0.039	1.032
60	-0.870	1.181	-0.018	1.047
70	-0.877	1.163	0.015	1.027
80	-0.946	1.217	0.016	1.049
100	-0.893	1.195	0.038	1.078
120	-0.808	1.134	0.047	1.382

model, Δg_{gpm}, is removed from the free-air anomalies, Δg_{fa}. Next, Stokes' formula with a Wong-Gore modified Stokes kernel is applied to the reduced gravity anomalies. At last, the geoid, N, is obtained by restoring the effect of the global geopotential model, N_{gpm}. The calculated geoid may be given mathematically as

$$N = \frac{R}{4\pi G} \iint_{\sigma} (\Delta g_{\mathrm{fa}} - \Delta g_{\mathrm{gpm}}) S^{\tau}(\psi) \mathrm{d}\sigma + N_{\mathrm{gpm}}, \tag{1}$$

where $S^{\tau}(\psi)$ is the Wong-Gore modified Stokes' function with cutoff degree τ, G is the average value for the gravity and R is the Earth's radius.

Several different geoids were calculated by changing the truncation degree, by using either EGM96 or EIGEN-2S as reference model and by either including or not including NIMA gravity data in the geoid determination.

6 Test procedure

The test procedure of the EIGEN-2S and EGM96 model is divided into two steps. In step one, the spherical harmonic expansion of the reference model is truncated at several different degrees, indicated by N_{max} in Tables 3 and 4.

In the second step, gravity data from the NKG data base, Arctic Gravity Project, NIMA and KMS02 data set are used in the calculation of the gravimetric geoids, together with a geopotential reference model. These gravimetric geoids are compared to GPS networks and MSSH-MDT data. Statistics are listed in Tables 5.

7 Results

7.1 NIMA gravity data

The inclusion of NIMA data in the geoid determination improves the fit to GPS and MSSH-MDT, see Table 5. The effect on the geoid is at the sub-meter level and is highly correlated to the location of the NIMA gravity points, see Fig. 5.

7.2 EIGEN-2S compared to EGM96

7.2.1 *Step 1*

The comparison of the global geopotential models (GGM) to the GPS network and the MSSH-MDT data are listed in Tables 3 and 4.

When compared to GPS and MSSH-MDT data the EIGEN-2S model gives slightly better results than EGM96 up to degree 50, see Tables 3 and 4.

It is also noticed that the EIGEN-2S has increasingly better fit up to around degree 70. Above that degree the fit decreases. This indicates a lack of information in the higher degrees of the model.

7.2.2 *Step 2*

Several different geoids were calculated by varying the cutoff degree τ in Eq. (1) and the truncation degree N_{max} of the global geopotential model (GGM). A few of these geoids are listed in Table 5. The table shows, using the EGM96 as reference model, that a gravimetric geoid gives generally better fit than geoids using EIGEN-2S. This is to be expected since the full EGM96 model contain far more information than the EIGEN-2S. This is also the case, however, not that significant, for truncation degrees down to 50. For lower degrees the EIGEN-2S gives better results, which is to be expected based on Tables 3 and 4.

The results indicate, using the EIGEN-2S model, that truncation degree $N_{\mathrm{max}} = 120$ give only slightly better fit than a truncation degree of 70, see Table 5.

Table 5. The subscript indicates which global geopotential model has been used. The EIGEN-2S model is indicated by the abbreviation *E2S*, the truncation degree, N_{max}, used, is given in parenthesis and the superscript $-$ indicates that no NIMA data has been used in the geoid calculation. The geoid, N^{NAT}_{EGM96}, is the Tallin geoid generated by Forsberg (2000). All values in m

Name	τ	GPS		MSSH-MDT	
		Mean	SD	Mean	SD
N^{NAT}_{EGM96}	36	-0.538	0.121	-0.077	0.203
N_{EGM96}	36	-0.479	0.248	0.059	0.207
$N_{EGM96(40)}$	36	-0.548	0.314	0.022	0.337
$N_{EGM96(70)}$	36	-0.467	0.252	0.057	0.240
$N_{E2S(40)}$	36	-0.557	0.323	0.009	0.329
$N_{E2S(70)}$	36	-0.530	0.257	0.047	0.250
$N_{E2S(120)}$	36	-0.535	0.257	0.052	0.238
N^{-}_{EGM96}	36	-0.480	0.247	0.059	0.209
$N^{-}_{E2S(70)}$	36	-0.531	0.257	0.048	0.251
$N^{-}_{E2S(120)}$	36	-0.535	0.257	0.052	0.241

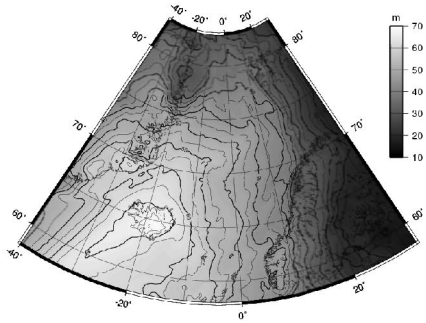


Fig. 3. The geoid with τ=36 and EGM96 (N_{max}=360). Illuminated from south; contour interval 2.5 m

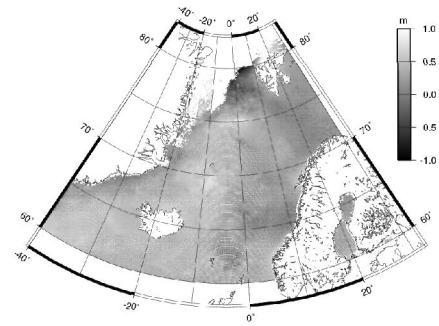


Fig. 4. The difference between the geoid [with τ=36 and EGM96 (N_{max}=360) as reference model and MSSH-MDT

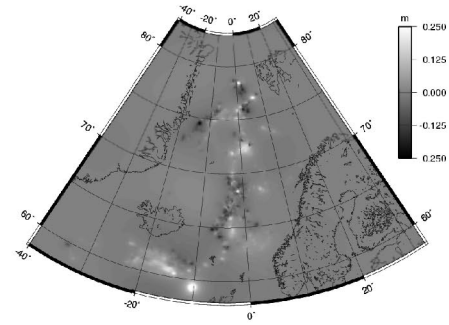


Fig. 5. The effect on the geoid due to the inclusion of NIMA data

8 Conclusions

Tables 3, 4 and 5 indicate that the EIGEN-2S reference model may be an improvement compared to the EGM96 up to degree 50, however, gravimetric geoids using the full EGM96 model as reference model generally give better results both over the Nordic countries and in the North Atlantic region.

By including the NIMA data set in the geoid determination the fit to GPS and MSSH-MDT data is slightly improved.

References

Andersen, O. (2003). Personal communication.

Andersen, O. B. and Knudsen, P. (1998). Global marine gravity field from the ERS-1 and GEOSAT geodetic mission altimetry. *J. Geophys. Res.*, **103**(C4), 8129.

Forsberg, R. (2000). New geoid from Greeland to the Ural Mountains. Joint Baltic and Nordic Gravity and Geoid Meeting, Tallin, Estonia.

Kenyon, S. (2003). Personal communication.

Lemoine, F., Smith, D., Pavlis, E., Pavlis, N., Klosko, S., Chinn, D., Torrence, M., Williamson, R., Cox, C., Rachlin, K., and Wang, Y. (1996). The development of the NASA/GSFC and DMA joint geopotential model. In *Proc. Int. Symp. of Gravity, Geoid and Marine Geodesy (GRAGEOMAR)*.

Reigber, C., Schwintzer, P., Neumayer, H., Barthelmes, F., Koenig, R., Förste, C., Balmino, G., Biancale, R., Lemoine, J.-M., Loyer, S., Bruinsma, S., Perosanz, F., and Fayard, T. (accepted Jan. 2003). The champ-only earth gravity field model. *Advances in Space Research*.

Computation of a geopotential model from GOCE data using fast spherical collocation - A simulation study

C.C. Tschernig[1], **D.N. Arabelos**[2]

[1] Department of Geophysics, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen Oe, Denmark
[2] Department of Geodesy and Surveying, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Abstract. Using a realistic orbit for GOCE, IAG SC7 has created a one month gravity gradient dataset from EGM96 to degree 300, with gradients referring to an instrument frame aligned with the velocity vector and the z-axis in the plane formed by this vector and the position vector. From the second order derivative of the potential V_{zz} we subtracted the contribution of EGM96 to degree 24. The resulted (noise free) data set was used to predict radial gravity gradient values in a $0.5°$ grid, covering the area of the Earth from $-83°$to $+ 83°$ latitude using local Least-Squares Collocation (LSC). The standard deviation of differences between predicted gridded values and values computed from EGM96 (degree 24 - 300) was between 1.0 and 0.5mE (Eötvös unit = E, 1 E=10^{-9}s^{-2}). Correlated noise with a 3 mEU standard deviation and a $35°$ correlation distance was added to the simulated data and the gridding was repeated. The formal LSC error-estimates were 2 mE. This was confirmed by comparing radial derivatives from EGM96 with the values predicted from the data with noise. The simulated data sets were used to generate spherical harmonic coefficients of the gravity potential to degree 300 using Fast Spherical Collocation (FSC), with a global covariance function. Both, a grid of noise-free data and a grid obtained from the data with noise were used. Both results agreed with EGM96 within the error-bounds of the FSC estimate.

Key words. GOCE mission · Geopotential model · Fast spherical collocation

1 Introduction

Let us denote by V the gravity potential of the Earth (without centrifugal part) and by T the anomalous potential obtained by subtracting a reference potential from V. Both functions may be approximated arbitrarily well as the sum of a series in solid spherical harmonics with coefficients C_{ij}

$$V(r,\varphi,\lambda) = \frac{GM}{a}\sum_{i=0}^{n}\left(\frac{a}{r}\right)^{i+1}\sum_{j=-i}^{i} C_{ij}Y_{ij}(\varphi,\lambda),$$

where r is the distance from the origin, φ, λ the latitude and the longitude respectively, GM the product of the gravitational constant and the mass of the Earth and Y_{ij} are the surface spherical harmonics. n is the maximal degree, in this paper generally equal to 300.

The ESA satellite GOCE (Gravity Field and Steady-State Ocean Circulation Explorer Mission, ESA, 1999) is expected to be launched in 2006. It will carry a gradiometer which will measure the 6 gravity gradients, $V_{xx}, V_{xy}, V_{yy}, V_{xz}, V_{yz}, V_{zz}$, given in the instrument frame with the x-axis aligned with the velocity vector, the z-axis in the plane formed by the velocity and the position vector. The y-axis completes the orthogonal triad. The goal of the mission is to determine geoid heights and gravity anomalies with an error below 0.01m and 1.0mGal, respectivelyr, at length scales of less than 100 km.

Many global and regional simulations (see e.g. Arabelos and Tscherning, 1995) have been carried out, initially using preliminary information available from ESA. Meanwhile the orbit parameters have been fixed, and more realistic simulations can be made. In order to enable the comparison between different methods IAG Special Commission 7 has issued a 1 month dataset with a 5s sampling, totally 718401 points with 6 gravity gradient values given in the above mentioned instrument frame. EGM96 to degree 300 (Lemoine et al., 1998) was used to generate the gradients.

Least-squares collocation may, as demonstrated in Tscherning (2001), be used to predict spherical harmonic coefficients. The use of the method for this purpose requires that a globally distributed set of data is used. Furthermore the optimal use of LSC require that systems of equations with as many unknowns as observations are solved, a procedure which obviously is not feasible today.

Fortunately the method of Fast Spherical Collocation, FSC, (Sansò and Tscherning, 2003) enables us to bypass this problem. But data gridded equidistantly in longitude and at a distance from the origin, which is constant for each parallel must be used. The method has also limitations with respect to which

data-types may be used. For gravity gradients the radial derivative, T_{rr} may be used, for example. In the following we will describe how such values were obtained from the "clean" SC7 data and from data to which correlated noise has been added. Finally we will present the results of the application of FSC to the two interpolated (gridded) data sets.

2 Prediction of gridded T_{rr} data

LSC with regionally distributed data may be used to interpolate/predict values of the gravity gradients with a limited effort. Other studies (Tscherning, 2003) show that excellent results may be obtained using T_{zz} to predict T_{rr} just using data in a neighbourhood which size will depend on the quality of the reference potential used to define T. This is due to the closeness of the radius vector to the instrument $z-$ axis. We will here use EGM96 to degree 24 as our reference potential. Using such a reference potential we will have a correlation distance of 1.5° at mean satellite altitude (255km), see Figure 1.

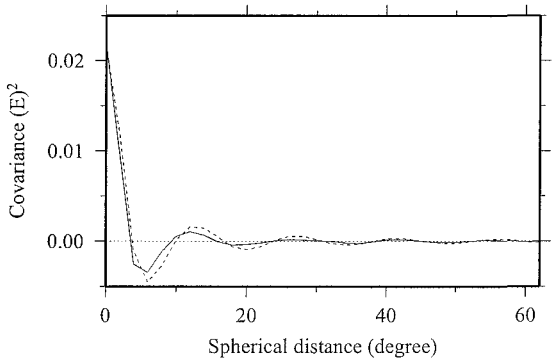


Fig. 1. Covariance function of second order radial derivative of T at 255 km altitude: solid line=from degree-variances to degree 720, dashed line=from analytic expression

Consequently, a region with at least a 1.5° border should be used. We have used 2.5° in latitude and longitude, i.e. the border is smaller at numerically high latitudes, but here will the data density be very high due to the 96° inclination of GOCE. In order to limit the administration of the regional solutions, we decided to divide the Earth in $20° \times 20°$ equal angular blocks, i.e. totally $9 \times 18 = 162$ regions.

LSC permits us to use covariance models fitted to local statistical characteristics. We decided only partly to use this possibility. In fact we simply scaled a global covariance function (Fig. 1) with respect to the regional variance. The global model was determined in two steps. First a covariance function of T_{rr} was determined for 255km altitude. It was determined as a Legendre-series with EGM96 error degree-variances from degree 2 to 24, EGM96 degree-variances from degree 25 - 360 and GPM98A (Wenzel, 1998) degree-variances from degree 361 - 720. This function was then fitted with an analytic expression using the GRAVSOFT (Tscherning et al., 1992) program COVFIT (Knudsen, 1987). This program fits a closed expression to the values derived from the finite Legendre series.

Using the GRAVSOFT program GEOCOL, 162 jobs were run (1 for each of the regions), where values in a 0.5° grid at 255km altitude as well as error-estimates were computed. The numerically maximal latitude was 83.25°.

Initially the noise-free SC7 T_{zz} data were used to predict the values which were compared to values computed from EGM96 (degree 24-300). An agreement was found with a mean value of zero and a standard deviation of typically 1 mE.

The error-characteristics of the GOCE gradiometer measurements will first be known after the launch of the satellite. However, power-spectral-density models of the expected noise are developed in (SID, 2000). We decided to use the same model as also used in earlier studies (Bouman et al., 2003). After a preprocesing removing contingent global biases, the data will have a correlated noise with a 3mE standard deviation and a 35° correlation distance.

The gridding the computations were then repeated with noise (see 2) added to the simulated data. The standard deviation between values calculated from EGM96 and predicted values now were in the range $1.5-2.0$mE, which corresponds well to a prediction error of 2.0mE estimated by LSC.

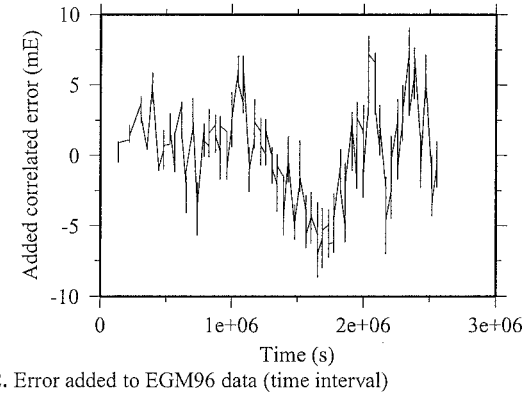


Fig. 2. Error added to EGM96 data (time interval)

In Figure 3 the mean and the standard deviation of the differences between "observed" and "predicted" T_{rr} as a function of latitude is shown. From this figure it is evident that the error is large at the block-boundaries and as it expected, increases very much in the area right outside the data area.

3 Estimation of spherical harmonic coefficients using FSC

The gridded data could directly be used by the program SPHGRID, which used a covariance function represented by the sum of a Legendre series to degree 720. The error-estimates were for all parallels fixed to 2 mE.

The standard deviation (unit less) per degree of the differences between the predicted coefficients and EGM96 coefficients (red curve) as well as the mean collocation error-estimates per degree of the predicted coefficients (blue curve) are shown in Figures 4 and 5. In the same figures the square

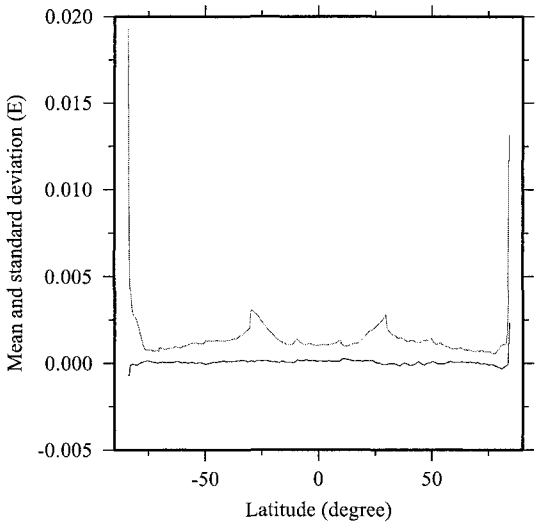


Fig. 3. Mean and standard deviation of the differences between observed and predicted T_{rr}, as a function of latitude. Black=Mean, Red=Standard deviation. The "tops" at latitudes 30° and −30° are caused by using a too small overlap zone at these degrees

roots of potential degree-variances of EGM96 are shown (error degree-variances from degree 2 to 24, degree-variances from degree 25 to 300).

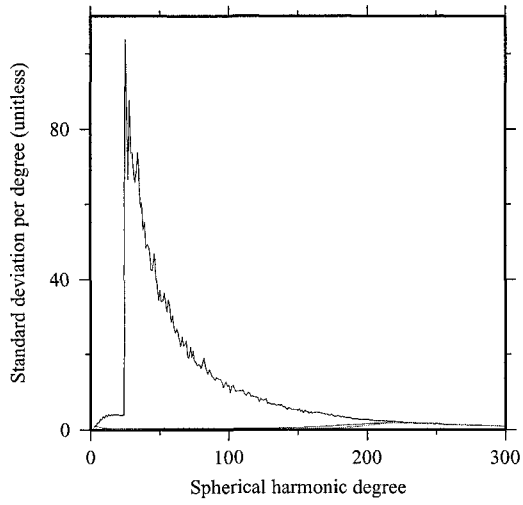


Fig. 4. Results of FSC prediction of spherical harmonics. Black line= Degree variances, Red line=Standard deviation per degree of the differences observed - predicted, Blue line=Mean collocation error-estimates of the predicted coefficients per degree

We see that the crossing point, where the signal and the noise standard deviation becomes equal are approximately at degree 200 as also found in (ESA, 1999, Fig. 4.1).

Figure 5 furthermore show the influence of aliasing. Recall that the simulated data are computed using a spherical harmonic series with coefficients of degree 25 - 300, while the covariance function has non-zero degree- variances from degree 2 to 720. Consequently the predicted coefficients with

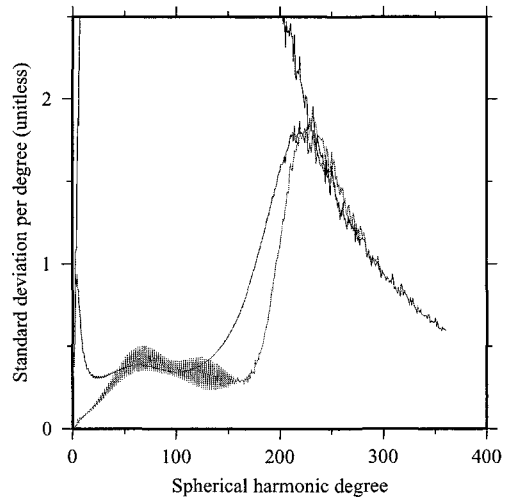


Fig. 5. Blow up of bottom part of Figure 4. Same colours used as in Fig. 4

the degree between 2 and 24 will not be zero. However the root-mean square difference (red in the figure) is still much smaller than the estimated error (blue). Also in the interval from degree 150 to 210 is the estimated error larger than the actual error. A possible interpretation of this is that LSC predicts that aliasing from degrees 301 - 720 will influence the result.

Figure 6 shows the error-estimates as a func- tion of the order for the coefficients of degree 150. The influence of the polar gap is clearly seen (cf, Tscherning, 2001a).

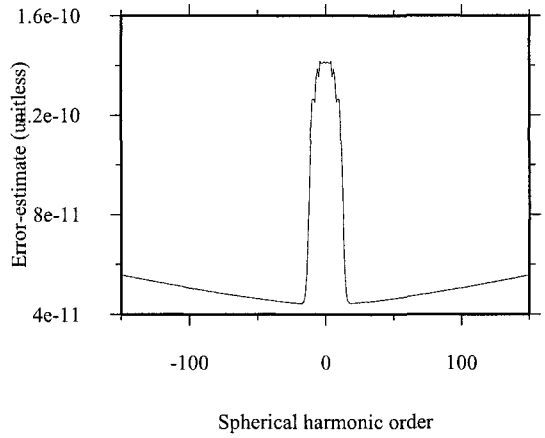


Fig. 6. Error estimates for degree 150 as a function of the order. Note the influence of the polar gaps

4 Conclusion

Regional LSC has been used to predict values of gravity gradients (here T_{rr}) in a grid. The grid values were used as input data to Fast Spherical Collocation to determine estimates of

spherical harmonic coefficients as well as error-estimates of these quantities.

It is possible to use predicted gridded values of other gravity gradients such as $T_{\varphi\varphi}$ and $T_{\lambda\lambda}$, the second order derivatives with respect to latitude and longitude. While T_{rr} is "close" to T_{zz} these quantities are, however, (depending on position), not necessarily close to the observables T_{xx} and T_{yy}. Thus combining all quantities in a gridding procedure may as demonstrated in (Tscherning, 2003) give excellent results.

Further investigations will be executed in order to see whether it is possible to improve the estimates of the coefficients using these additional data types and by adding data in the polar areas.

There are also other problems which require further investigations. We have used regional covariance models derived in a simple manner by scaling a global model. Here actual estimates of the regional covariance function could be used as discussed in (Arabelos and Tscherning, 2003).

There is also an inconsistency between the covariance models and the simulated data used. The data only contain coefficients with degrees between 25 and 300, while the covariance functions contain degrees from 2 to 720. This means that we are not seeing the full influence of aliasing.

References

Arabelos, D. and Tscherning, C.C., Regional recovery of the gravity field from SGG and Gravity Vector data using collocation. *J. Geophys. Res.*, *100*, (B11):22009–22015, 1995.

Arabelos, D. and Tscherning, C.C., Globally covering a-priori regional gravity covariance models. *Advances in Geosciences*, *1*, 143–147, 2003.

Bouman, J., Koop, R., Sneuw, N. and Tscherning, C.C., Terrestrial Data for Calibration and Error Assessment of GOCE Gravity Gradients. Presented at EGS/EUG/AGU meeting, Nice, 2003.

ESA: Gravity Field and Steady-State Ocean Circulation Mission, ESA SP-1233, 1999.

Knudsen, P., Estimation and Modelling of the Local Empirical Covariance Function using gravity and satellite altimeter data. *Bull. Géod.*, *61*, 145–160, 1987.

Lemoine, F.G., Kenyon, S.C., Factor,J.K., Trimmer, R.G., Pavlis, N.K., Chinn, D.S., Cox, C.M., Klosko, S.M., Luthcke, S.B., Torrence, M.H., Wang, Y.M., Williamson, R.G., Pavlis, E.C., Rapp, R.H., and Olson, T.R., The Development of the Joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM96, NASA/TP-1998-206861, Goddard Space Flight Center, Greenbelt, MD, July, 1998.

Sansò, F. and C.C.Tscherning: Fast Spherical Collocation - Theory and Examples. *J. Geod.*, *77*, 101–112, 2003.

SID: GOCE End to End Performance Analysis. Final Report ESTEC Contract no. 12735/98/NL/GD. SID, 2000.

Tscherning, C.C.: Computation of spherical harmonic coefficients and their error estimates using Least Squares Collocation. J. of Geodesy, Vol. 75, 14–18, 2001.

Tscherning, C.C., Covering the GOCE mission polar gaps using gradients and ground gravity. Proc. International GOCE User Workshop, ESTEC, 23-24 April 2001, 105–110. ESA WPP-188, 2001a.

Tscherning, C.C., Testing gridding and filtering of GOCE gradiometer data by Least-Squares Collocation using simulated data. Paper prepared for IUGG2003, 2003.

Tscherning, C.C., Forsberg, R. and P.Knudsen, P., The GRAVSOFT package for geoid determination. Proc. 1. Continental Workshop on the Geoid in Europe, Prague, May 1992, 327–334, Research Institute of Geodesy, Topography and Cartography, Prague, 1992.

Wenzel, H.-G., Ultra high degree geopotential models GPM98A and GPM98B to degree 1800 Proc. Joint Meeting Int. Gravity Commission and Int. Geoid Commission, Budapest, 10-14 March. Rep 98:4, Finnish Geodetic Institute, Helsinki, 71–80, 1998.

Improved Determination of Sea Surface Heights Close to the Australian Coast from ERS-2 Satellite Radar Altimetry

X. Deng
School of Engineering, The University of Newcastle,
University Drive, Callaghan, NSW 2308, Australia, Email: xiaoli.deng@newcastle.edu.au

W.E. Featherstone
Western Australian Centre for Geodesy, Curtin University of Technology,
GPO Box U1987, Perth, WA 6845, Australia, Email: W.Featherstone@curtin.edu.au

Abstract. The homogeneous and repeated data coverage over coastal regions from satellite radar altimetry is one important data source for oceanographic and geodetic applications. However, the sea surface heights (SSH) extracted from the altimeter data are often in error close to the coast, due in part to the complex nature of echoes returned from rapidly varying coastal topographic surfaces (both land and sea) and the generally rougher sea state. This paper presents improved SSH results derived from ERS-2 altimeter waveform data (two cycles of 35-day repeat orbit, March to May 1999) near the Australian coast using a coastal retracking system. This system was developed based upon a systematic and comprehensive analysis of return waveforms. Central to the system is the use of two retracking techniques: the iterative least squares fitting and the threshold retracking algorithms. Using the AUSGeoid98 geoid grid as a quasi-independent ground reference and comparing with a broad contaminated distance of ~10 km obtained from a previous study before retracking, the use of the retracking system is able to reduce this contaminated distance to ~5 km. However, improved SSH data cannot be recovered by waveform retracking less than ~5 km from the coastline due to predominant land returns in the waveform range window.

Keywords. Satellite radar altimeter waveforms, coastal waveform retracking system, improved sea surface height.

1 Introduction

Although many previous studies (e.g., Mantripp 1996, Brooks *et al.* 1997) have reported that sea surface height (SSH) data are in error in coastal regions, the potential of altimeter waveform data to be used in these regions to obtain improved SSHs has been the subject of comparatively little study. Therefore, this study is important in that the improved altimeter SSHs can be obtained and will be able to provide reliable and uniform data sources for geodetic and oceanographic applications in coastal regions.

So far, considering the existing retracking algorithms, two approaches have been applied to retracking coastal waveforms, i.e., the fitting and threshold techniques. The first is to fit an analytical waveform model to the observed waveform data for determining the more accurate range estimate (e.g., Anzenhofer *et al.* 2000). The fitting function used from the open literature is based upon the five-parameter function developed by Martin *et al.* (1983), choosing either the linear trailing edge or the exponential trailing edge (Anzenhofer *et al.* 2000). An alternative method used by Brooks *et al.* (1997) is a modified threshold retracking technique (cf. Wingham *et al.* 1986, Davis 1997, Zwally 1996), which retracks the ocean return from the mixed waveforms of water and land close to the coastline. Only a single retracker is used from these previous studies.

A comprehensive analysis of waveform shapes around the Australian coast (Deng *et al.* 2001, 2002) led to the need to develop a coastal waveform retracking system. When reprocessing and analysing waveform data, it was found that using only a single retracker in coastal regions limits the precision of recovered SSHs due to the disagreement between the theoretical and contaminated waveforms.

Therefore, it has become clear that no single retracking algorithm can always deal with the diverse waveforms in coastal regions. A coastal waveform retracking system that consists of different retrack-

ers (e.g., the parametric fitting and the threshold algorithms) is thus necessary when attempting to extract the precise SSH from retracked range measurements. The primary idea of the retracking system is that retrackers must handle most of the coastal waveforms, and can supplement each other. It is also significant that the system can correctly categorise the waveforms, so that they can be retracked using an appropriate retracker.

The principal objective of this study is to develop a coastal waveform retracking system, and to determine to what extent the SSH data can be improved by this system in coastal regions. As part of the development and implementation of the retracking system, it is important to apply the system to a coastal region and to assess the quality of the retracked altimeter-derived SSH data. With this consideration, two cycles of ERS-2 20 Hz waveform data (42 and 43, March to May 1999) will be retracked close (0-30 km) to the Australian coastline using the system.

To obtain a general idea of the impact of retracking on the altimeter-derived SSH, the SSH before and after retracking will be compared with an external data source. The external quasi-independent reference used is the 2′×2′ AUSGeoid98 geoid model (Featherstone *et al.* 2001). This is chosen because it is the most recent and precise model readily available in Australian coastal regions at the time of this study.

It is acknowledged that the instantaneous SSH is not equivalent to the geoid. The difference between them includes the time variation (e.g., tides) and dynamic sea surface topography (SST). The mean dynamic sea-surface height released by Levitus *et al.* (1977) is generally used to remove the quasi-independent SST from altimeter-derived SSHs. Since there were no data sources near the Australian coast, the SST values at this area are based on an interpolated procedure from a 1°×1° grid (Hwang, 2003). Thus, applying such a SST correction may cause an additional error to the SSH in this study, which might be larger than the errors of other geophysical and environmental corrections. Therefore, the SST is not applied to the SSH data in Australian coastal regions in this study. The verification of any improvement in the SSH will only be seen in the standard deviations.

2 Coastal Waveform Retracking System Development

Waveform retracking in coastal regions aims to determine the offset between the pre-designed tracking gate for the on-satellite tracking software and a known, fixed, instrument-independent position (i.e., the mid-point) on the leading edge of the waveform, and correct the satellite range measurement accordingly. The algorithms used in the system involve the fitting and modified threshold retracking techniques, but uses six retrackers.

Ideally, a physically realistic model should be used to retrack coastal waveforms, because the waveform measured by a radar altimeter links not only to the geometric surface of the ocean, but also to oceanic geophysical features. An appropriate model will provide a correct understanding of the relationship between the geophysical parameters of interest and the radar return from coastal regions.

With this consideration, an ocean model, which is the slightly modified Brown (1977) model and includes a linear ocean surface, Earth curvature and the off-nadir angle, was first used to retrack ocean-like waveforms. To retrack narrow and high-power waveforms reflected from calm coastal or inland water surfaces, the threshold retracking technique (Wingham *et al.* 1986) with a 50% threshold level was also used. It was found from retracked results in this study that both retrackers could not deal with some waveforms in close proximity to the land. Fitting an ocean model to non-ocean-like waveforms causes failure of the fitting algorithm. This led to the use of other retrackers of the five- and nine-parameter fitting functions (Anzenhofer *et al.* 2000, Zwally 1996) because of their ability to retrack non-ocean waveforms.

The 50% threshold level cannot always give improved results, particularly in areas closer to the coastline. To validate the threshold level of the retracking algorithm in these areas, a statistical analysis has been performed to suggest the threshold levels based upon the quality of the retracked SSH data. The evaluation turned out a detailed selection of the threshold levels (given below).

When retracking individual waveforms in coastal regions using all possible retrackers, the results show improved SSHs, but biases exist among different algorithms. Thus, an algorithm assessment has been implemented to analyse and estimate these biases (given below). All these investigations motivate the development of the coastal retracking system.

Finally, a coastal retracking system (Fig. 1) has been developed to retrack altimeter waveforms in Australian coastal regions, which involves waveform data editing, waveform classification, iterative fitting procedure and threshold technique.

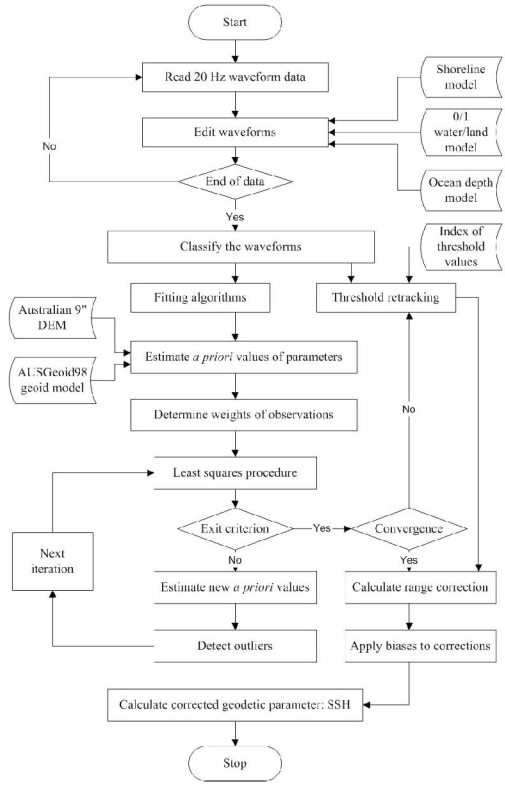


Fig. 1 Block diagram of the coastal retracking system

2.1 Selection of the Threshold Level

When using the threshold retracking algorithm, it is important to select an appropriate threshold level. There are two concerns related to this selection. The first is that threshold values are higher than the noise in the early gates. The second is that the values are sufficiently lower than the sought-after ocean returns. Therefore, considering the shape variations of the waveform in the proximity of the coasts, the threshold levels used in the coastal retracking system differ from the method of Brooks *et al.* (1997) and previously published methods (e.g., Zwally 1996, Davis 1995, Davis 1997).

A varying threshold level is used in this study, which sets different threshold levels depending on whether or not the waveform is contaminated. If the waveform classification indicates a distorted waveform, the threshold level is set to 30% (or 20%) of the waveform amplitude; otherwise 50% is set. Compared with the waveform shape, the distance from the coast is not significant in this criterion. It links the threshold level to the contaminated waveform, selecting a threshold level automatically.

In order to evaluate this threshold level, one cycle of 20 Hz ERS-2 waveform data (March to April 1999) was used in areas around the Australian coast. Since the waveform contamination strongly depends on both the distance from the coast and land surface topography along the coastline, computation is conducted in six 5 km-wide bands around the coast. Waveforms have been threshold retracked using different threshold levels.

For each along-track altimeter-derived SSH (both before and after threshold retracking), the gridded AUSGeoid98 geoid height (as a quasi-independent ground truth) is bispline-interpolated to the longitude and latitude of the point. The difference between the SSH and geoid height can be obtained by subtracting the interpolated geoid height from the corresponding SSH. Then, the descriptive statistics of these differences are computed.

Figure 2 shows positive standard deviations (STD) of the mean differences in a distance band of 5-10 km from Australia's coastline. As can be seen from the ordinate scale in Fig .2, the STDs decrease with increasing the distance from the coasts. It is therefore evident that the SSH is improved after retracking no matter which threshold level is used. However, the 50% threshold level does not show improved results in all areas, while the results suggest that the varying threshold level can be the best one.

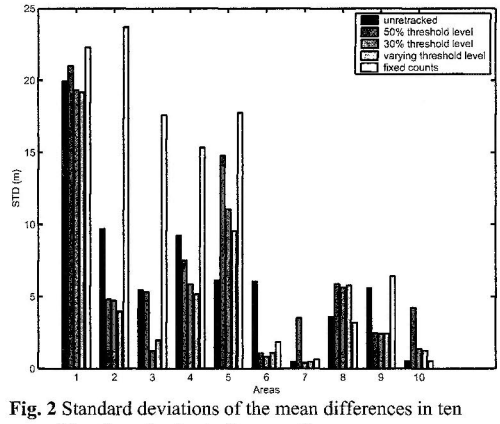


Fig. 2 Standard deviations of the mean differences in ten areas 5 km from the Australian coastline.

2.2 Biases among Retracking Algorithms

The biases among different retracking algorithms have been reported by Anzenhofer *et al.* (2000). In

order to investigate the biases, waveform data over an open ocean area (50-350 km from the coastline, -30°S<φ<-15°S, 105°E<λ<120°E) are retracked using the system. There are 54500 ocean waveforms in this area. Of these data, there are 233 unusual waveforms (~0.4%), which present no leading edge information in the range window. These unusual waveforms were deleted before retracking. Of the remaining data, the fitting algorithm is able to retrack ~98% of the waveforms.

Using cycles 42 and 43 of ERS-2 waveform data, the biases have been estimated. The bias between two fitting models is small (~ -0.8 cm) and almost statistically insignificant. This is because they are both derived from the Brown (1977) model. However, it is significant to find bias values of ~57 cm and ~56 cm between the 50% threshold algorithm and the five-parameter model and the ocean model, respectively. These values indicate that the range retracked by the fitting model is ~56.5 cm longer than that retracked by the 50% threshold retracking algorithm. These biases have been removed from the retracking system.

3 Results and Discussion

3.1 Results of Waveform Categories

Following the steps of the retracking block diagram in Fig. 1, two cycles (42 and 43) of ERS-2 20 Hz waveform data have been retracked over a small part of the Australian coast (0-30 km from the coastline). The results from waveform categorising show that most contaminated waveforms (~80%) have the ocean-like shape and fewer of them (~12%) have non-ocean-like shapes. These non-ocean-like waveforms are generally found much closer to the coastline (<10 km), but they can also occur anywhere within the study distance. In particular high-peaked waveforms were found at a distance of 0-20 km from the coastline. This indicates further the fact that using only a single retracker (e.g., the five-parameter model) cannot reprocess all waveforms in coastal regions. Other retrackers are more significant in such areas to remove errors caused by waveform contamination.

3.2 Analysis of Single Track Retracked Results

The SSHs along two ground tracks (Fig. 3) 21364 and 21085, both leaving land to water, will be analysed in this section. In order to evaluate how the waveform is contaminated near the coast and improved by the retracking techniques, the AUSGeoid98 geoid height has been interpolated at each location along the tracks. Both profiles of the SSHs before and after retracking and the geoid height are plotted versus along track distances in Fig. 4.

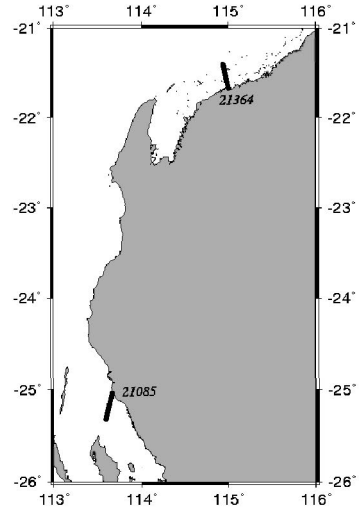


Fig. 3 Two ERS-2 ground tracks off the northwest Australian coast

As stated, it is acknowledged that the geoid height does not equal the SSH measured by the altimeter. The difference is evident in Fig. 4. Therefore, the difference between the SSH and geoid height is not taken as an indicator for the purpose of comparison or analysis. Instead, the profile's trends of the SSH and geoid or the along-track slopes will be considered as a more appropriate indicator to examine the waveform contamination and the improvement of the retracking procedure.

Comparison of the geoid heights with unretracked SSHs from both tracks shows that the latter is deviating up to ~5 m respect to the AUSGeoid98 geoid (Fig. 4). As can be seen from Fig. 4, the typical near-land effects on unretracked SSH profiles begins at the along-track distance of ~20 km and ~17.5 km from the coastline for the track 21364 (top of Fig. 4) and track 21085 (bottom of Fig. 4), respectively. When leaving the coastline from land to water, unretracked SSHs decrease and then increase gradually within ~10 km for both tracks. In contrast to the change of unretracked SSHs, the geoid heights show a near-linear trend over the whole distance.

According to the shape of the unretracked SSH profiles, the along-track SSH and geoid gradients are computed. The results for the track 21085 show

that along-track geoid gradients computed from the AUSGeoid98 grid are small, 9.2-10.9 ppm when the groundtrack is 0-7.5 km and 7.6-19.5 km from the coastline. Corresponding to the same distances, the SSH gradients computed from unretracked SSHs are 975.6 ppm and -509.2 ppm, respectively. They are much larger than the maximum geoid gradient of ~150 ppm observed anywhere in Australia (cf. Featherstone *et al.* 2001). Beyond 19.5 km from the coastline, the AUSGeoid98 gradients and SSH gradients computed using unretracked SSHs show small values of 12.5 ppm and -6.4 ppm, respectively. These results, thus, suggest that unretracked SSHs are in error within an along-track distance of 0-19.5 km from the coastline.

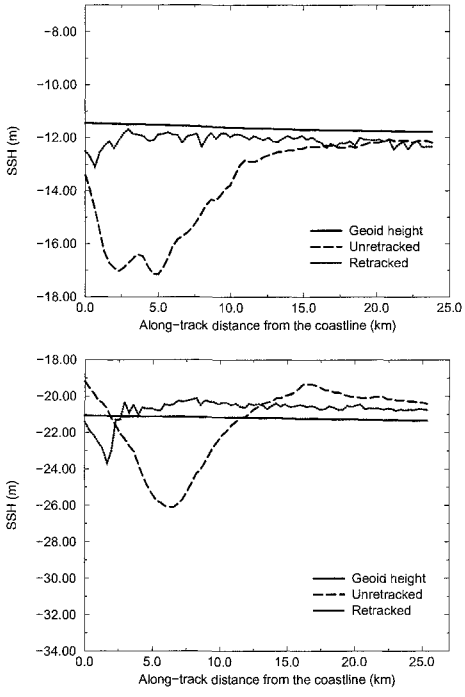


Fig. 4 Geoid height and SSH profiles (ERS-2, cycle 42) along ground track 21364 (top) and 21085 (bottom) before and after retracking using the coastal retracking system.

Comparing the SSH profiles before and after retracking in Fig. 4, the SSH profiles after retracking show a good agreement of the profile trend with the geoid-height profile at distances from ~2.5 km to ~25 km from the coastline. The contaminated distances are reduced to ~2.5 km. However, SSHs in this distance cannot be effectively improved by the waveform retracking system because of the increase of land-dominated returns.

It is also apparent from Fig. 4 that the retracked SSH profile becomes noisier as compared with the unretracked SSHs. The noise might be caused by, firstly, the shorter wavelength signal (i.e., temporal variations of the scattering surface that the on-board tracking algorithm cannot closely follow); secondly, unmodelled non-linear parameters in this study (e.g., the skewness). Thus, a low-pass filtering procedure is still essential when using 20 Hz SSH data after retracking for geophysical application in coastal regions.

3.3 Evaluation of SSH Data Using a Geoid

In order to assess the precision of the retracked data and to determine how the altimeter retracked SSH profiles represent the true sea surface, the descriptive statistics of the differences between SSH data (20 Hz) before and after retracking and the AUSGeoid98 has been computed. For the convenience of comparison, the STD (only positive value) of the differences have been plotted via six 5-km-wide distance bands in Figs. 5 and 6 for cycles 42 and 43. From Figs. 5 and 6, the STDs show large values for both SSH data before and after retracking. The reasons may be the temporal variations of the ocean surface, SST included in the SSHs, and other incorrect corrections, such as ocean tides and wet tropospheric range corrections. However, the retracked SSH data show a higher precision than the unretracked data.

For both cycles, Figs. 5 and 6 indicate that the STDs decrease when using retracked SSH data. The values of STD drop to ~1 m and significant improvement occurs in distances 0-15 km from the coastline. Beyond 15 km, there is no significant decrease in the value of the STD of differences. These results suggest that a more precise SSH profile can be obtained from retracked SSHs, especially 0-15 km from the coastline.

The results from Figs. 5 and 6 also show a better precision of the retracked SSH data in Australian coastal regions than unretracked SSH data. The STD of the differences after retracking drops to ~ ±1 m when the distance from the coastline is more than ~5 km. This indicates clearly that the retracked SSH can be determined as precisely as SSHs over the open oceans beyond ~5 km, while unretracked SSH data have the same precision after only ~15 km from the coastline. This means that these improved SSHs after retracking in coastal regions can be used with sufficient precision to

compute coastal geoidal undulations, coastal ocean circulation, gravity anomalies and ocean tides.

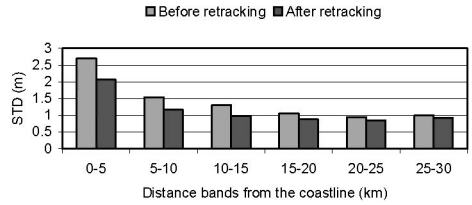


Fig. 5 STD of the mean difference between the SSHs and AUSGeoid heights before and after retracking in six 5 km-wide distance bands (cycle 42).

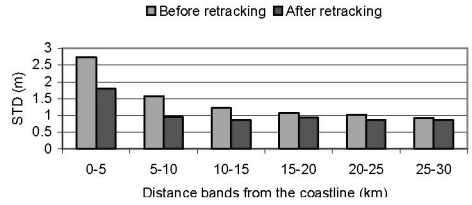


Fig. 6 STD of the mean difference between the SSHs and AUSGeoid heights before and after retracking in six 5 km wide distance bands (cycle 43).

5 Conclusions

The coastal waveform retracking system developed in this study has improved and extended altimeter-derived SSHs ~10 km shoreward in Australian coastal regions compared with unretracked SSH data, using AUSGeoid98 as a quasi-independent reference. The altimeter-derived SSH profiles are now more precise beyond ~5 km from the Australian coastline. These improved SSHs show the same precision level as those over open oceans, thus indicating a great potential to apply them to altimetric-geodetic applications in coastal regions. However, SSH profiles 0-5 km from the coastline still cannot be recovered by the waveform retracking procedure at the same precision level as SSH data beyond 5 km from the coastline. This is attributed to sever contamination from land returns and the rougher sea states.

It is necessary in the future for users to remove the quasi-independent SST from the SSH using a local model in this area, and for European Space Agency to overcome the problem in coastal regions.

Acknowledgments. This research is supported by an Australian International Postgraduate Research Scholarship (IPRS). The authors would like to thank ESA for kindly providing ERS-2 waveform data. We would also like to thank Professors P.A.M. Berry, C. Hwang, and M.G. Stewart for useful discussions.

References

Anzenhofer, M., C.K. Shum and M. Renstch (2000). Coastal Altimetry and Applications, Ohio State University Geodetic Science and Surveying Report.

Brooks, R.L., D.W. Lockwood and J.E. Lee (1997) *Land Effects on TOPEX Radar Altimeter Measurements on Pacific Rim Coastal Zones*, http://topex.wff.nasa.gov/does [on line].

Brown, G.S. (1977). The Average Impulse Response of a Rough Surface and Its Applications, *IEEE Transactions on Antennas and Propagation*, AP-25(1), pp. 67-74.

Chelton, D.B., J.C. Ries, B.J. Haines, L.-L. Fu and P.S. Callahan (2001). Satellite Altimetry, in: *Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications*, Fu, L.-L. and A. Cazenave, (eds.), Academic Press, San Diego, USA, pp. 1-132.

Davis, C.H. (1995). Growth of the Greenland Ice Sheet: A Performance Assessment of Altimeter Retracking Algorithms, *IEEE Transactions on Geoscience and Remote Sensing*, 33(5), pp. 1108-1116.

Davis, C.H. (1997). A Robust Threshold Retracking Algorithm for Measuring Ice-Sheet Surface Elevation Change from Satellite Radar Altimeter, *IEEE Transactions on Geoscience and Remote Sensing*, 35(4), pp. 974-979.

Deng, X., W.E. Featherstone, C. Hwang and P.A.M. Berry (2001). Retracking ERS-2 Radar Altimeter Waveform Data in Australia's Coastal Zone, *Proceedings of the IAG 2001 Scientific Assembly*, Budapest, Hungary, September [CD-ROM].

Deng, X., W.E. Featherstone, C. Hwang and P.A.M. Berry (2002). Estimation of Contamination of ERS-2 and POSEIDON Satellite Radar Altimetry Close to the Coasts of Australia, *Marine Geodesy*, 25(4), pp. 249-271.

Featherstone, W.E., J.F. Kirby, A.H.W. Kearsley, J.R. Gilliland, G.M. Johnston, J. Steed, R. Forsberg and M.G. Sideris (2001) The AUSGeoid98 Geoid Model of Australia: Data Treatment, Computations and Comparisons with GPS-Levelling Data, *Journal of Geodesy*, 75, pp. 313-330.

Hwang, C. (2003). *Personal Communication*, National Chiao-Tung University, Taiwan.

Levitus, S., G.I. Monterey and T. Boyer (1997) Seasonal Variability of Dynamic Height and Its Fourier Analysis, *NOAA NESDIS Atlas 15*, Government Printing Office, Washington, D.C., USA.

Martin, T.V., H.J. Zwally, A.C. Brenner and R.A. Bindschadler (1983). Analysis and Retracking of Continental Ice Sheet Radar Altimeter Waveforms, *Journal of Geophysical Research*, 88(C3), pp. 1608-1616.

Mantripp, D. (1996). Radar Altimetry, in: *The Determination of Geophysical Parameters from Space*, Fancey, N. E., Gardiner, I.D. and Gardiner, R.A. (eds.), Institute of Physics Publishing, London, UK, pp. 119-171.

Wingham, D.J., C.G. Rapley and H. Griffiths, (1986). New Techniques in Satellite Tracking Systems, *Proceedings of IGARSS'86 Symposium*, Zürich, September, pp. 1339-1344.

Zwally, H.J. (1996). The GSFC Retracking Algorithms, http://icesat4.gsfc.nasa.gov/, [on line].

A numerical study of solving the altimetry-gravimetry boundary value problem in coastal regions

R.S. Grebenitcharsky and M.G. Sideris
Department of Geomatics Engineering, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, Canada, T2N 1N4; Email: grebski@ucalgary.ca; Fax: +1-403-284-1980

Abstract. The paper presents numerical results for the solution of the altimetry-gravimetry boundary value problem with compatibility conditions along the coastline. The solution is based on the application of spherical Shannon wavelets based on the Abel-Poisson kernel and pseudo-diffferential operators. The orthogonality of Shannon wavelets can be used for the numerical solution overcoming some technical problems with generally non-orthogonal spherical wavelets.

A numerical experiment has been conducted in eastern Canada and the results have been compared with the geoid heights computed from gravity anomalies by applying 1D FFT spherical Stokes convolution with 50 km integration radius. The effect of compatibility conditions along the coastline on the solution are discussed and the following conclusions are drawn: spherical Shannon wavelets can be very successfully applied in the numerical solution of the AGBVP II and the smoothing effect of spherical wavelets is much more effective than the classical planar type of wavelets.

Keywords. Geoid, altimetry-gravimetry boundary value problems, satellite altimetry, gravity,wavelets

1 Introduction

Altimetry-gravimetry boundary value problems (AGBVPs) provide the frame in which the combination of different types of data is possible. Historically, the first type of AGBVPs appeared in Sansò (1981) and Arnold (1983), known later as AGBVP I. Another type of AGBVP formulated in Holota (1980) is used when gravity data are available at sea, known later as AGBVP II. Holota treated both problems in a linearized form in Holota (1983a, 1983b). Later, with the availability of GPS/levelling data, a new AGBVP appeared, called AGBPV III (Lehmann, 1999). All AGBVPs consist of two different types of geodetic boundary data sets, one on land and one at sea. They depend on the type of boundary surface and boundary conditions (measurements) used. In Sansò (1993), different types of geodetic boundary value problems (GBVPs) are described both on land and at sea. As mixed BVPs, all AGBVPs have problems on the boundaries between different types of data and to overcome these problems compatibility conditions along the coastline are necessary.

The role of the compatibility conditions for altimetry-gravimetry boundary-value problems is quite complicated, including the existence of the solution, smoothing boundary conditions, providing higher level of regularization and smoothness of the solution. They are usually used to ensure the well-posedness of the problem (Sacerdote and Sansò 1987, Svensson 1988, Lehmann 1999). The uniqueness and existence of the solution of AGBVPs depend upon the smoothness assumptions along the coastline.

Another use for compatibility conditions along the coastline is related to the regularization of the solution in the neighbourhood of the coastline. In reality, irregularities exist on the coastline because different types of data with different observation procedures, accuracy and resolution are used. According to Svensson (1988), unless the compatibility conditions are recognized, the mixed problems cannot be solved with a higher degree of smoothness. Additional constraints on the data for the regularization of the solution across the coastline were derived in Sacerdote and Sansò (1987). The regularization of data along the coastline will regularize and guarantee a reasonable smoothness of the solution. The addition of compatibility conditions along the coastline may cause the problem to become well posed, but it is questionable whether such conditions upon the data are realistic for practical use. To answer this question, three preliminary numerical experiments have been conducted; see Grebenitcharsky and Sideris (2001a), (2002b) and Grebenitcharsky et al. (2001). The first one investigates the effect of smoothness conditions applied to the same type of data on land and at sea – gravity anomalies from gravity and shipborne gravity measurements. The second one investigates the effect of smoothness conditions along the coastline applied to different types of data – GPS/leveling and gravity data on land *together* with altimetry and shipborne gravity data at sea. In both cases, the effect of smoothing resulted in a geoid change between -5 cm and +5 cm. This magnitude can be considered significant for cm-geoid determination. The third experiment (Grebenicharsky and Sideris, 2002b) was conducted to investigate data

smoothing and its effect on the geoid determination along a mountainous coastline. The effect of smoothness on the final geoid solution along the coastline was between –33.1 cm and 24.7 cm in the test area. The magnitude of the smoothing effect in rough terrain and complicated coastline is much greater than in flat areas. In all three experiments, the applied smoothing procedure using wavelet decomposition and reconstruction had an effect only in the coastal area but in the other regions located away from the coastline there were no changes in the boundary conditions (measurements) and final geoid.

These preliminary numerical results encourage a further theoretical investigation for finding solutions of AGBVPs based on the combination of spherical harmonics, spherical pseudo-differential operators, and spherical wavelets for solving differential equations; see Grebenitcharsky and Sideris (2002a).

Applying spherical harmonic expansion (spherical Fourier transform) for the definition of pseudo-differential operators (PDOs) will give us the so-called spherical PDOs in the finite spherical domain. They have been derived from the theory of singular integrals. For AGBVPs, PDOs can be applied not only for the reformulation of AGBVPs in a simple way but also for imposing compatibility conditions at the coastline. The theory of spherical PDOs has been described in Freeden and Windheuser (1997) and in Freeden et al. (1998). All PDOs are related to the idea of generalized functions, which in fact creates difficulties in the application of compatibility conditions along the coastline. At the same time, compatibility conditions have to be locally applied. In this case, the combination between spherical PDOs and spherical wavelets can help us to overcome this problem.

The application of multiresolution analysis (MRA) and wavelets with well known types of wavelet transforms is not suitable for the solution of AGBVPs because known wavelets for infinite (planar) domain are not harmonic functions. To solve the problem, new wavelet methods for approximating harmonic functions have been suggested in Freeden and Schreiner (1995). Spherical wavelets are harmonic and have good space and spectral localization properties, which are necessary for the application of compatibility (smoothness) conditions. As a spherical scaling function for the solution of AGBVPs, the most suitable kernel is the Abel-Poisson kernel because it is harmonic (Freeden and Windheuser, 1996). When the harmonic Legendre polynomials are used as base functions, the corresponding spherical scaling function is harmonic as well.

It is known that the Abel-Poison kernel produces nonorthogonal spherical wavelets. As a result, numerical difficulties can be expected and the application of a spherical multi-resolution analysis (MRA) will require an orthogonalizaton procedure. Another way to overcome this problem is to use H-Shannon spherical wavelets with the Abel-Poisson kernel as a scale function (Freeden and Michel, 2002). They are orthogonal and have local support, which makes them very useful for the local application of compatibility conditions along the coastline in the numerical solutions of AGBVPs.

The main task for this paper it to give a procedure and conduct a numerical experiment for the numerical solution of the AGBVP II, applying the compatibility conditions along the coastline as detail wavelet coefficients restrictions. This is possible because: 1) the compatibility conditions applied on the sphere are equivalent to the requirement that the data, and the first and second order gradients be consistent along the coastline; 2) of the properties of wavelets to behave as multiscale differential operators.

2 The AGBVP II and the role of compatibility conditions along the coastline – spherical approximation

According to Sansó (1993) and Lehmann (1999) the AGBVP II in spherical approximation, will have the form:

$$\begin{cases} \nabla^2 T = 0 \quad in \quad \Omega \quad (outside \quad the \quad sphere \; \delta\Omega) \\ L_S(T) = -\dfrac{\partial T}{\partial r} = \delta g_S \qquad (known \quad \delta\Omega_S) \\ L_L(T) = -\dfrac{\partial T}{\partial r} - \dfrac{2}{R} T = \Delta g_L \qquad (unknown \quad \delta\Omega_L) \\ T = O(|\xi|^{-2-n}) \quad |\xi| = r \to \infty \end{cases} \tag{1}$$

where $\delta\Omega$ is the sphere as the boundary of the space Ω outside the sphere; the indices L and S denote the land and sea part of the boundary surface, ∇^2 is the Laplace operator, T is the disturbing potential, L_S, L_L are linear functionals of T representing the boundary conditions (measurements) at sea and on land, δg_S is the gravity disturbance at sea, Δg_L is the gravity anomaly on land, and R is the Earth radius, $-n-2$ is the order of the decreasing disturbing potential and it is related to the degree of the geopotential model used. The last equation represents the requirement that the disturbing potential is regular at infinity.

Introducing PDOs on the sphere as in Svensson (1988) we can reformulate the AGBVP II and establish the following compatibility conditions:

$$\begin{cases} AT = \Delta g_L \qquad on \quad \delta\Omega_L \\ (A+2)(A+3)AT = (-\nabla^2 - 2)\delta g_S \quad at \quad \delta\Omega_S, \\ (A+2)SL\Delta g_L + (A+2)SM(-\nabla^2 - 2)\delta g_S \\ + (A+2)T_1 = 0 \qquad at \quad coastline \end{cases} \tag{2}$$

where A is a PDO, defined in Eq. 4, L and M are extensions from the land and the sea part of the sphere to the space outside the sphere (see Grebenitcharsky and Sideris, 2002a), S is Stokes's operator and T_1 is the first degree term of T. On the sphere $\delta\Omega$ with radius R, the Laplacean will have the following representation in spherical coordinates:

$$\nabla^2 = R^{-2}(\partial^2/\partial\varphi^2 + \tan\varphi\,\partial/\partial\varphi + \cos^{-2}\varphi\,\partial^2/\partial\lambda^2) \quad (3)$$

The Laplacean is itself a pseudo-differential operator and it appears in the pseudo-differential operators when imposing compatibility conditions at the coastline. The PDOs used in the new formulation of AGBVP have the following form:

$$\begin{aligned} A &= (-\nabla^2 + 1/(4R^2))^{1/2} - 3/(2R), \\ C &= (-\nabla^2 + 1/(4R^2))^{1/2} + 3/(2R), \\ AC &= CA = -\nabla^2 - 2/R^2, \quad C = A + 3/R. \end{aligned} \quad (4)$$

It can be shown that the inverse spherical PDOs A^{-1}, C^{-1}, $(AC)^{-1}$, and different combinations of them, do exist and they have certain physical meanings. For example, A^{-1} is Stokes's operator and $(I-C)^{-1}$ is Hotine's operator. The operator C itself is the gradient operator. Taking into account that the inverse PDOs exist, AGBVPs can be redefined, compatibility conditions can be established, and the solution can be found numerically. Compatibility conditions in this form have mainly theoretical meaning and it is necessary to discuss their relationship to the regularity of the data along the coastline.

We can prove that on a unit sphere the compatibility conditions (in spherical approximation) correspond to the following conditions:

$$\begin{cases} \delta g_S = \Delta g_L + 2T & at \quad coastline \\ C\delta g_S = C\Delta g_L - 4A^{-1}T & at \quad coastline \\ \nabla^2\delta g_S = \nabla^2\Delta g_L & at \quad coastline \end{cases} \quad (5)$$

Let us start with the second condition in the form

$$AC\delta g_S = AC\Delta g_L - 4T \quad (6)$$

and assume that L and M are two complementary operators such as $L+M=P$, where the operator P could be simply interpreted as merging data on land and at sea. Like in the planar case in Grebenitcharsky and Sideris (2001a), we can show in a similar way that

$$\{I - I\}\{(A+2)SL\Delta g_L + (A+2)SMAC\delta g_S\} = 0 \quad (7)$$

or

$$\{(A+2)SL\Delta g_L + (A+2)SMAC\delta g_S\} = -(A+2)T_1 \quad (8)$$

which is exactly the compatibility condition at the coastline.

Now, to improve the regularity of the data (and the regularity of the solution as well) for the spherical approximation, we can impose the above conditions on the data, first and second order gradients and even for higher orders gradients. This can be done with the application of wavelets as multiscale differential operators (Mallat, 1998). In our numerical solution, we are going to use spherical H-Shannon wavelets based on the Abel-Poisson kernel and we will limit the detail coefficients to zero value.

3 Numerical solution of AGBVP II with compatibility (smoothing) conditions along the coastline

The fact that the boundary surface at sea is known from the altimetry data and having shipborne gravity data allows us to transform gravity disturbances at sea to gravity anomalies using the following equation:

$$\Delta g_S = \delta g_S + \frac{2G}{R} N_{alt} \quad (9)$$

where G is mean gravity and N_{alt} is the geoid height from altimetry and corrected for sea surface topography (SST).

Having gravity anomalies both at sea and on land we can apply the Stokes kernel in terms of spherical PDOs together with spherical wavelets (Freeden and Windheuser, 1997). After the wavelet decomposition step, we restrict the detail coefficients to a certain value (zero for the numerical experiment) for every level of decomposition, which is equivalent to the restriction of gradients of first or higher order. As a result, the data along the coastline will be smoothed and the reconstruction step with the convolution between the dual wavelet and the Stokes's kernel will give the solution for the smoothed data along the coastline. The computational procedure to get the numerical solution is the following:

Step 1: Transform gravity disturbances at sea to gravity anomalies using geoid heights from the altimetry data. Remove the contribution of the geopotential model (EGM96) up to degree n_{GPM}=360. As a result we obtain residual gravity anomalies Δg_{GPM}.

Step 2: Grid the residual gravity anomalies for land and sea separately. Merge both sets of data using a topographic model to identify grid points at sea and on land.

Step 3: Determine the low-pass filter L_D and high-pass filter H_D for the decomposition step.

3.1 Determine the scale function and corresponding wavelet function using the formulas for H-Shannon spherical wavelets given by Freeden and Michel (2002):

The scale function Φ_j for *j-level* of decomposition is

$$\Phi_j(\psi) = \sum_{n=n_{GPM}}^{N_j} h^n \frac{2n+1}{4\pi} P_n(\cos(\psi)) \quad (10)$$

The wavelet function Ψ_j for *j-level* of decomposition is

$$\psi_j(\psi) = h^{N_j+1} \frac{2N_j+3}{4\pi} P_{N_j+1}(\cos(\psi)) \quad (11)$$

if $h=r/R=1$ all data are on the sphere without upward continuation, N_j is the degree of Legendre polynomials corresponding to *j-level*, n_{GPM} is the degree of the used reference geopotential model.

3.2 Determine both decomposition filters based on the scale and the wavelet function applying the built-in function in MATLAB.

Step 4: Determine the low pass filter L_R and high pass filter H_R for the decomposition step.

4.1 Determine the Stokes kernel as a Legendre polynomial series according to Freeden and Windheuser (1997) as

$$K_\Lambda(\psi) = \sum_{n=0}^{N_j} \frac{2n+1}{4\pi} \hat{\Lambda}(n) P_n(\cos(\psi)) \quad (12)$$

where the spherical symbol for the Stokes operator is $\hat{\Lambda}(n) = 1/(n-1)$ and $n \neq 1$. For other types of data the only change is the spherical symbol – for example, for gravity disturbances it would be $\hat{\Lambda}(n) = 1/(n+1)$.

4.2 Determine the corresponding low-pass and high-pass filters for the reconstruction step by convolving the ones from the decomposition step with the Stokes kernel. In terms of formulas, we have

$$L_R(\psi) = L_D(\psi) * K_\Lambda(\psi), \quad H_R(\psi) = H_D(\psi) * K_\Lambda(\psi) \quad (13)$$

Step 5: Perform one level of wavelet decomposition and reconstruction using the corresponding low-pass and high-pass filters after smoothing the data along the coastline by limiting detail coefficients up to zero. The magnitude of detail coefficients away of the coastline are zero after the decomposition step.

5.1 Perform the decomposition step using the L_D and H_D filters – the results are approximation plus detail coefficients. The coastline is expected to be detected in the detail coefficients because of existing discrepancies between data along the coastline.

5.2 Apply the compatibility (smoothing) conditions by setting the detail coefficients to zero value.

5.3 Perform the reconstruction step using the L_R and H_R filters. The result is the approximated (within a scale constant) geoid height residuals $\Delta\tilde{N}_{GPM}$ smoothed along the coastline.

Step 6: Compute the final smoothed geoid heights on the grid points for the area under study.

6.1 Scale the approximated geoid height residuals to have the correct geoid height residuals ΔN_{GPM} on the grid

$$\Delta N_{GPM} = \frac{R\,\Delta\varphi\,\Delta\lambda\cos(\varphi)}{\gamma_0} \Delta\tilde{N}_{GPM}. \quad (14)$$

6.2 Obtain final smoothed geoid heights by restoring the contribution of the geopotential model.

$$N = N_{GPM} + \Delta N_{GPM}. \quad (15)$$

4 Numerical experiment, data and results

A numerical experiment has been made to investigate the proposed procedure for the solution of AGBVPII. The area under study is $43° \le \varphi \le 57°$ and $298° \le \lambda \le 312°$ (Newfoundland, Eastern Canada, Fig. 1). The following data were used: free air gravity anomalies on land from the Geological Survey of Canada, shipborne free air gravity anomalies at sea from the Geological Survey of Canada and the GEOSAT geodetic mission (GM) satellite altimetry data at sea from NOAA. The altimetry data close to the coastline are extrapolated from the available data at sea. The EGM96 geopotential model was used as reference field. To show more details along the coastline, some results are presented in a smaller window (see Fig 1).

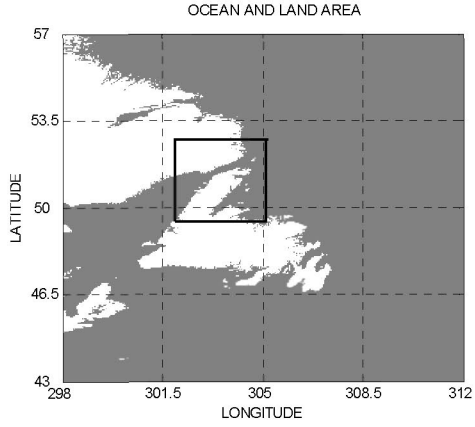


Fig. 1: Area under investigation

The effect of smoothing on the measurements can be seen in Fig 2. The original gravity anomalies, the smoothed gravity anomalies and the differences between both show the ability of spherical wavelets to detect irregularities in data along the coastline. The greatest differences were along the coastline. The effect of smoothing on the measurements was between -45.5 and +61.4 mGal (Table 1). The smoothing effect is mostly along the coastline; data further from the

A comparison with one level of decomposition and reconstruction with classical planar Daubechies 8 wavelets (see Fig. 3) showed much better smoothing properties for spherical wavelets. In this case, the data along the coastline have been changed a little bit but still the discrepancies exist.

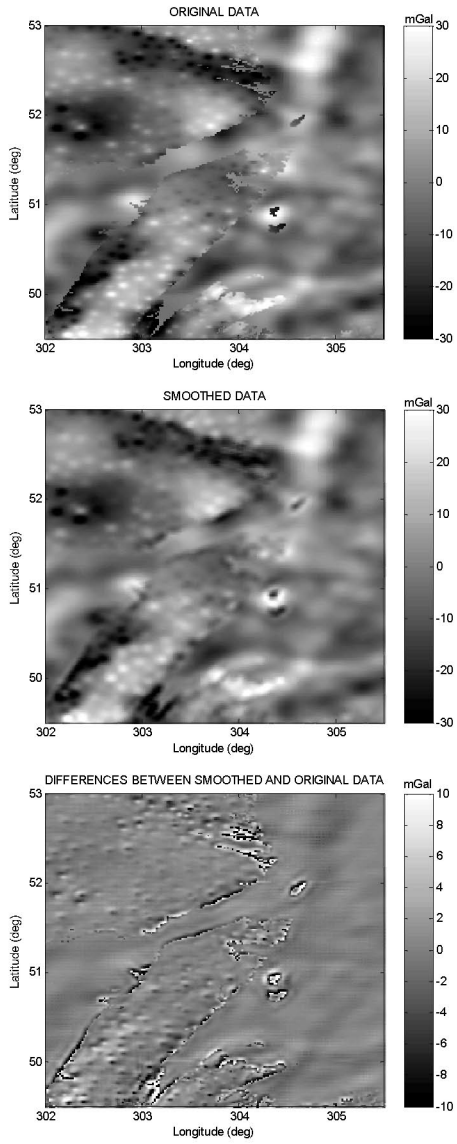


Fig. 2: Effect of smoothing with spherical wavelets

Table 1. Residual gravity anomaly differences (smoothed minus original) Unit: [mGal].

max	min	mean	RMS	STD
61.4	-45.5	0.0	1.6	1.6

The main advantage of the proposed numerical procedure is that the geoid heights are available immediately after the reconstruction procedure. Also, for other types of data, it is necessary to only change the spherical symbol in the kernel represented by a Legendre polynomial series.

To validate the numerical solution, the statistics of the differences in geoid heights between the proposed numerical solution and the 1D FFT spherical Stokes convolution with 50 km integration radius (Sideris, 1999) were computed and are shown in Table 2.

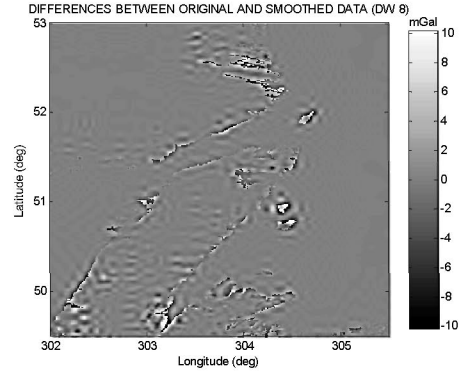


Fig. 3: Effect of smoothing with Daubechies8 wavelets

Table 2. Geoid height differences with respect to 1D FFT Unit: [m].

max	min	mean	RMS	STD
0.50	-0.30	0.0	0.10	0.10

The statistics show that the proposed numerical solution is close to the one by 1D FFT (with an RMS difference of 10 cm) and that further investigation of the effect of smoothing on the final geoid heights is necessary.

Fig. 4 shows the smoothing effect of compatibility conditions on the final geoid heights. Again, all changes were along the coastline and between -3.4 and +3.6 cm (see Table 3).

Table 3. Differences between the geoids from smoothed and original data. Unit: [m].

max	min	mean	RMS	STD
0.036	-0.034	0.0	0.001	0.001

A smoothing effect of 4 cm is significant for cm-geoid determination. It agrees with what was seen in the preliminary study of the smoothing effect. Taking into account the fact that the area under study is flat, the effect of smoothing for a mountainous and very complicated coastline is expected to be much greater.

5 Conclusions and recommendations

The following conclusions can be drawn:

- Spherical Shannon wavelets with Abel-Poisson kernel can be very successfully applied in the numerical

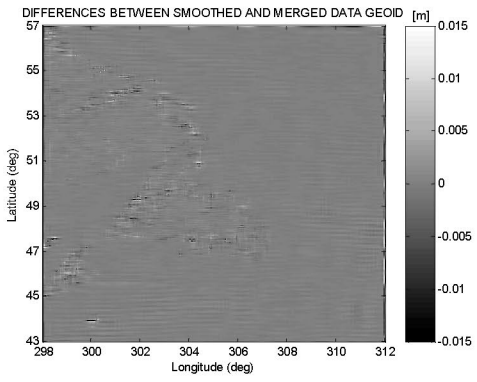


Fig. 4: Effect of smoothing on the final geoid heights

solution of the AGBVP II. From a theoretical point of view the solution for the disturbing potential will be harmonic because the used wavelets are harmonic. From a practical point of view, the orthogonality of wavelets helps us overcome a lot of technical complications. The very narrow local support of scale and wavelet function allows the computation to be done in a local area using a grid on the sphere.

- The geoid heights can be immediately computed in the reconstruction step. The application for other types of data requires only changes in the spherical symbols used for different functionals of the disturbing potential.

- The smoothing effect of spherical wavelets is much more effective, compared to the classical planar type of wavelets and it is concentrated along the coastline. No additional information about the coastline is necessary for the procedure to detect and smooth automatically the data irregularities along the coastline.

- The smoothing effect on the coastline has a magnitude of -4 to 4 cm, which corresponds to our previous investigations and is significant for cm-geoid determination.

- The proposed solution has to be applied in a mountainous and complicated coastline like the western coast of Canada and US. The expected smoothing effect under these conditions should be much larger than the one in flat area we studied here.

Acknowledgements: This research has been funded by grants from the Natural Sciences and Engineering Research Council of Canada and by the GEOIDE Network of Centers of Excellence.

References

Arnold, K(1983) The mixed boundary value problem of geodesy. Gerlands Beitr Geophys 92, 391-406.

Freeden, W. and Michel, V. (2002) Orthogonal Non-bandlimited wavelets on the sphere, University of Kaiserslautern, Laboratory of Technomathematics, 21-22.

Freeden, W. and Schreiner, F. (1995) New wavelet methods for approximating harmonic functions. In: Geodetic theory today (F. Sanso, ed.) International Association of Geodesy Symposia, 114, Springer, 112-121.

Freeden, W. and Windheuser U. (1996) Spherical wavelet transform and its discretization, Adv. Comp. Math., 5, 51-94.

Freeden, W. and Windheuser, U. (1997) Combined spherical harmonic and wavelet expansion – A future concept in Earth's gravitational determination, Appl.,Comp., Harm. Anal., 4, 1-37.

Freeden, W. Gervens, T. and Schreiner, M. (1998) Constructive approximation on the sphere, with applications to Geomathematics, Oxford Science Publications.

Grebenitcharsky, R. and Sideris, M.G. (2001a) Altimetry-gravimetry boundary value problems with smoothness conditions in coastal regions, in Tziavos I. N. and Barzaghi R. (ed.) EGS 2001-G7 Session "Regional and local gravity field approximation", Nice, France, 25-30 March, 2001, Bulletin of IGeS N. 13, special issue, 121-132.

Grebenitcharsky. R. and Sideris, M.G. (2001b) A Comparison of Different Solution Methods for Altimetry-Gravimetry Boundary Value Problems Using Smoothing Conditions Along the Coastline, Presented at the 2001 International Association of Geodesy Scientific Assembly, Budapest, Hungary, September 2-7.

Grebenitcharsky, R., Vergos, G.S., and Sideris, M.G. (2001) Combination of gravity, altimetry and GPS/Leveling data for the numerical solution of the altimetry-gravimetry boundary value problems, In: Adam J., Schwarz K.,P. (eds) Vistas for geodesy in the new millennium, IAG Scientific Assembly, Budapest, Hungary, September 2-7, 150-155.

Grebenitcharsky R., Sideris M.G. (2002a) Application of spherical pseudo-differential operators and spherical wavelets for numerical solutions of the fixed altimetry-gravimetry boundary value problem, presented at "V Hotine-Marussi symposium on Mathematical geodesy", Matera, Italy, June 17-21, 2002 (in print).

Holota P (1980) The Altimetry-gravimetry Boundary Value problem, presented at the International Scientific Conference of Sec.6 Intercosmos, Albena.

Holota, P. (1983a) The altimetry gravimetry boundary value problem I: linearization, Friedrich's inequality. Boll Geod Sci Afini, 42, 14-32.

Holota, P. (1983b) The altimetry gravimetry boundary value problem II: weak solution, V-ellipticity. Boll Geod Sci Afini, 42, 70-84.

Lehmann, R. (1999) Boundary-value problems in the complex world of geodetic measurements, *J of Geodesy 73* pp. 491-500.

Mallat, S. (1998) A wavelet tour of signal processing. Academic press, USA.

Sacerdote. F. and Sansò, F. (1987) Further remarks on the altimetry-gravimetry problems. Bull Geod, 61, 183-201.

Sansò F. (1981) Recent Advances in the theory of the geodetic boundary value problem, Review of Geoph. and Space phys., 19/3

Sansò, F. (1993). Theory of geodetic B.V.P.s applied to the analysis of altimeter data. In: Rummel R, Sanso F. (eds) Satellite altimetry in geodesy and oceanography. Lecture Notes in Earth Sciences, vol 50. Springer, Berlin Heidelberg New York.

Sideris, M.G., (1999) Geoid Determination by FFT techniques – International school for the determination and use of the geoid, International Geoid Service, DIIAR-Politecnico di Milano, 165-229

Svensson L (1988) Some remarks on the altimetry-gravimetry problem. Manuscr Geod, 13, 63-74.

Improved High Resolution Altimetric Gravity Field Mapping (KMS2002 Global Marine Gravity Field)

Ole B. Andersen and P. Knudsen
Kort- og Matrikelstyrelsen, Geodesy, DK-2400 Copenhagen, Denmark. Email: oa@kms.dk, pk@kms.dk

Ron Trimmer
National Geospatial-Intelligence Agency, St. Louis, USA. Email: trimmerr@nga.mil

Abstract. A series of KMS high resolution near global gravity fields from altimetry have been released during the past 7 years. This presentation focuses on describing the improvements leading to the release of the new KMS2002 field.
Significant improvements have been achieved through careful fine-tuning of the procedure and parameters used for the derivation of gravity from satellite altimetry.
Particularly emphasis has been made in improving the gravity field determination at medium wavelengths ranging between 100-500 km, and at shorter wavelength in coastal regions. Emphasis was made in recovering the gravity field in lakes, which have normally been deleted from global marine gravity field maps. Extensive comparisons carried out at the National Imagery and mapping Agency is also presented to document the improvements made.

Keywords. Satellite Altimetry, Gravity, Arctic Ocean, inland lakes.

1 Introduction

Two satellites have operated in shorter non-repeating geodetic mission configurations. GEOSAT geodetic mission altimetry contains data from the first 18 months of the GEOSAT satellite mission (1985-1986) where the satellite was operated in a non-repeating orbit. This yielded a very dense, though not completely homogeneous coverage of observations. ERS-1 completed its 336-day geodetic mission in March 1995. Altimetry from this mission covers all oceans between -82° and 82° latitudes and provides a very dense (8.3 km) and homogeneous coverage

The altimetric sea surface height observation h can be described in terms of a geoid N and sea surface topography ξ. The geoid can be separated into a long wavelength geoid N_{REF}, (the EGM96) and residuals ΔN to this. Similarly the sea surface topography can be described in terms of a mean dynamic topography (ξ_{MDT}) and a time varying or dynamic sea surface topography ($\xi(t)$) (i.e., tides). Finally e is the additional error. The observed sea surface height can then be written like:

$$h = N + \xi + e = N_{REF} + \Delta N + \xi_{MDT} + \xi(t) + e$$

Detailed descriptions of the derivation of the gravity field from satellite altimetry with specific emphasis on the KMS gravity field can be found in Andersen and Knudsen, (1998, 2000); Andersen et al. (1996, 1999) The focus in this paper is to describe a few selected steps leading to the derivation of the KMS2002 Particularly emphasis in the derivation of the KMS2002 field has been made in the following steps

- Improving the quality and coverage of the altimetric height observations.
- Improving the gravity field mapping in coastal regions.
- Improving the reference long wavelength geoid model at medium wavelengths ranging between 100-500 km
- Improving the modelling of regional in-homogenities in the gravity field.
- Including gravity field mapping in lakes.

The following chapter will focus on these different details in the derivation of the gravity field. The subsequent chapters will deal with an extensive comparison with marine gravity observations followed by a summary.

2 The KMS20002 Global Marine Altimetric Gravity Field.

2.1 Improving the quality and coverage of the sea surface height observations.

The altimetric data used for the KMS99 and previous gravity fields were replaced for the derivation of the KMS2002 gravity field. Data from the Radar Altimeter Data Archive from DEOS TU-Delft (Scharroo et al, 1999) are now used for KMS2002. The DEOS altimetry archive contains data with the latest and highest quality within orbit modelling and path corrections.

Table 1. Orbit models and path corrections applied to the altimetric observations for the KMS99 and the KMS2002 altimetric observations. All altimetry observations from ERS and GEOSAT respectively used the same corrections.

	KMS2002 (ERS/GSAT)	KMS99 (ERS/GSAT)
Orbit	DGM-E04 / JGM-3	JGM-3 / JGM-3
Dry Troposp.	Standard / ECMWF	Standard / NCEP
Wet Troposp.	Model / Model	MWR / Model
Ionosphere	Bent / IRI95	Bent / IRI95
Inv Bar Corr	Std local / Std local	Std local / Std local
Ocean Tide	GOT00.2X ext	AG95
Load Tide	GOT00.2X ext	AG95
Pole Tide	Standard	Standard
Sea State Bias	BM3 / Std	None.

Altimetry from the non-repeating geodetic mission of ERS-1 and GEOSAT were used along with averaged repeat altimetric observations from the ERS-1 and ERS-2. Furthermore altimetry from repeat 60 through 63 of the ERS-2 mission was also included because the satellite drifted 8 km away from the average ground track during these repeats.

The major impact arises from the use of the Delft DGM-E04 orbit model for ERS data and the usage of an ECMWF based model correction for the wet troposphere correction for the ERS observations. The Micro Wave Radiometer (MWR) correction normally applied on ERS degrades close to the coast and resulted in loss of data closer than 20 km to the coast. Similarly the ocean tide model for both ERS and GEOSAT was upgraded from the AG95 (Shum et al., 1997) to the newer GOT00.2 model (Ray, 1999). The effect on the derived gravity in coastal regions is shown in the following section.

2.2 Improved gravity field modelling in coastal regions.

The gravity field in coastal regions are mainly improved from changing the wet troposphere and the tidal correction. The replacement of the MWR correction of the ERS-1 data increased the number of available observations close to the coast, but does not increase the actual quality of the observations.

By implementing a new tidal model the "gravity" signal to "oceanic" noise is improved because the ocean tidal contribution to the sea surface height $\xi(t)$ is better modelled. This is particularly important for the non-repeating geodetic mission observations where $\xi(t)$ can not be averaged in time and this way reduced.

The tidal correction for GEOSAT and ERS were replaced by a slightly modified version of the GOT00.2 ocean tide model by Ray, (1999) called GOT00.2X. This model is extended in coverage towards and onto the coast. This is done to avoid the problem of having altimetric observations without tidal correction, which would otherwise result in a large erroneous gravity field signal.

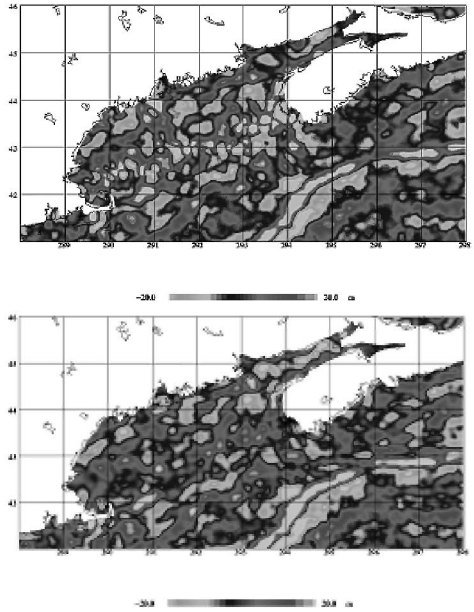


Figure 1. Derived altimetric gravity field in the Gulf of Maine. KMS99 based on AG99 (top), and KMS2002 based on GOT00.2X (bottom). The values are in mGal.

Figure 1 shows the KMS99 and KMS2002 gravity fields derived using the AG95 and GOT00.2X ocean tide models respectively. Notice the difference in the

mount of the Gulf of Maine where the KMS99 gravity field were in error because of a less accurate tidal model. Figure 2 shows the location of marine gravity observations and Table 2 summarizes the comparison with marine data in the Gulf of Maine region, which clearly indicate the gravity field improving from improving the ocean tide correction.

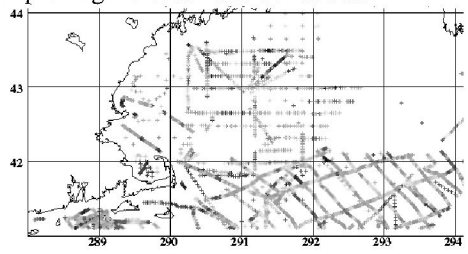


Figure 2. Location of 67789 Marina observations in the Gulf of Maine.

Table 2. Comparison with 67789 marine observations in the Gulf of Maine.

67789 obs	Mean	Std Dev.	Max abs
KMS02	0.68	5.38	29.90
KMS99	0.72	6.75	52.03

2.3 Modelling the medium wavelength in the gravity field.

The KMS gravity field is derived in cells of 2° latitude by 10° longitude and a remove-restore technique with respect to EGM96 (Lemoine et al., 1997). Therefore errors in the geoid model of wavelength longer than roughly 150-200 km will not be captured in the KMS approach (Andersen and Knudsen, 1998). Particularly at high latitude the EGM96 is known to have errors of more than 1 meter. (Reigber et al, 2002).

Assuming the mean dynamic topography ξ_{MDT} in the Arctic Ocean and around Antarctica is very small the altimetric height observations h can be used to derive long wavelength geoid and associated gravity field corrections. This was done by initially averaging all altimetric observations within 1° cells (to average out the dynamic topography).

By transforming the coordinate into polar coordinates 2D Fast Fourier Techniques was used to derive the gravity and geoid corrections to EGM96 for the regions above the 65° parallels. The resulting geoid and gravity field was compared with results obtained using 1-D spherical FFT approach yielding nearly identical results. The derived altimetric long wavelength geoid adjustment to the EGM96 geoid model is shown in the Arctic Ocean in Figure 3.

As the assumption of small mean dynamic topography is not valid at mid and low latitudes, so another approach was taken here. The observed heights were averaged in 0.25° blocks and associated gravity and geoid anomalies were estimated in 4° latitude by 20° longitude cells using same approach as used normally to derive KMS gravity fields. Finally all patches of corrections were combined to give complete coverage.

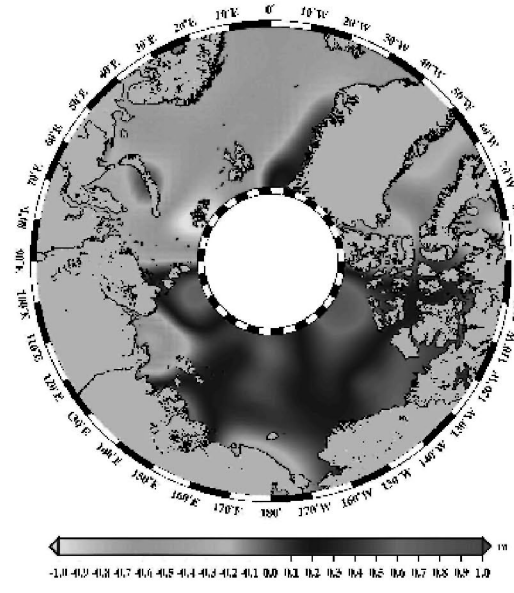


Figure 3. Derived long wavelength geoid adjustment to the EGM96 geoid model in the Arctic Ocean.

2.4 Modelling regional characteristics of the gravity field.

The gravity field relative to EGM96 varies greatly in variance regionally. This can be taken into account in the collocation-interpolation of the residual geoid anomalies by modelling this in the applied covariance functions.

In the derivation of the KMS99 and KMS2002 gravity fields, the interpolation is carried out using a modified second order Gauss Markov Covariance function like

$$c(r) = C_0(1+\frac{r}{\alpha})e^{(-r/\alpha)} + D_0(1+\frac{r}{\beta})e^{(-r/\beta)}$$

where the signal variance is C_o. The correlation length of the gravity signal is α. The additional terms with a variance of D_0 was introduced to model the remaining surface variability, which may still be present after the crossover adjustment. For observations on different tracks D_0 is fixed at zero yielding the normal second order Gauss Markov covariance expression. The correlation length β for the residual surface variability was empirically determined to 100 km. The variance of the error D_0 is computed from the altimetry (Andersen et al., 1999).

The signal variance C_o and the correlation length of the gravity signal (α) will vary spatially. In Figure 3 the spatial distribution of the signal and variance and the correlation length have been computed globally in 2° cells from empirically computed covariance functions. The large gravity structures associated with i.e., Hawaiian chain is clearly seen. Relatively short correlation lengths are seen along the shallow spreading ridges.

Attempts to include these spatially varying fields for variance for the signal variance C_o and correlation length degraded the derived gravity field slightly. This is most likely due to a combination of several reasons. Firstly, because the parameters C_0 and D_0 were both computed from altimetry and they were spatially correlated. Secondly, the spatial resolution of 2° for these fields may not be adequate enough for our purpose. Consequently, for KMS2002 a decision was made to keep the signal variance C_0 fixed at $(0.2\ m)^2$, and the correlation length α fixed at 18 km until further investigations has been carried out.

2.4 Gravity field modelling in Lakes.

Altimetric sea surface height observations are normally removed by an automatic editing procedure, if they depart from the reference geoid model by more than a certain threshold. This means that important altimetric observations from lakes like the Great lakes in the US (elevation roughly 150 meters) and the Caspian Sea (-34 meters) are discarded. Due to the land mask controlling the gain used by the ERS-1 satellite these observations were automatically removed and cannot be restored, but for GEOSAT the observations can actually be restored. The GEOSAT residual geoid height data ΔN *is* shown in Figure 4 and Figure 5. Notice the very large gradient in the Caspian Sea (10 meters) indicating the geoid problems in EGM96 in this region. Subsequently a gravity correction for the elevation of the lake was applied.

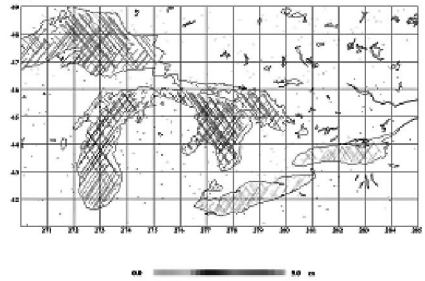


Figure 4. Residual geoid height data ΔN in the Great Lakes (only Geosat data available). An offset of 150 meters was applied to the height. The scale is in meters(+/- 5 meters).

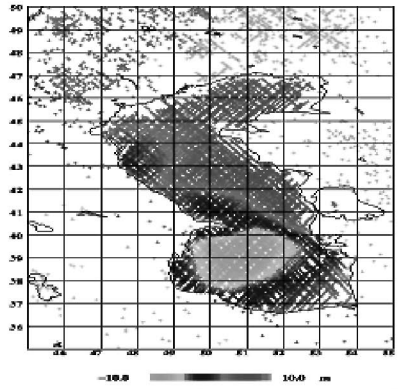


Figure 5. Residual geoid height data ΔN in the Caspian Sea (only Geosat data available). An offset of –34 meters was applied. The scale is in meters.

3 Comparison with Marine Gravity Observations

The comparison presented is a continuation of the work described in Trimmer et al, (2002). It compares the KMS02, KMS99, the GSFC00 field by (Wang, 2001) and the Sandwell and Smith (1997) version SSv8grav 2 with unclassified NIMA gravity observations in very difficult test region.

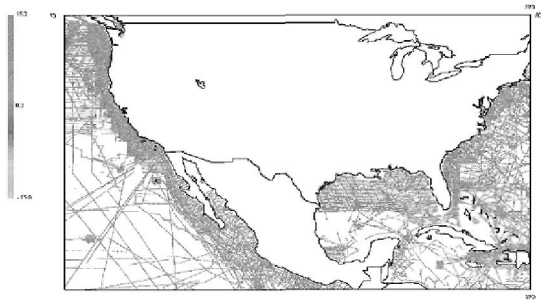


Figure 6. Differences between the KMS marine gravity field and the observed gravity from ship (color bar ranges between −15 mGal (blue) to 15 mGal (red))

Table 4. Comparison with 3,433,653 marine gravity observations. All values are in mGal.

	KMS2002	KMS99	GSFC	SS v8
Mean	.074	.074	-.302	-.600
Std Dev	5.781	6.108	6.050	7.340
Minimum	-360.293	-460.220	-458.150	-457.300
Maximum	196.980	211.787	140.350	139.108

Comparison with more than 3 million marine gravity observations is shown in Figure 6 with statistics presented in Table 4. The northwest part of the whole test area is designated as the Coastal Test Area. It is used to study regional error as a function of the distance to shore variable. The area is defined by the distance to shore values, which roughly covers the area within 100 km of shore along the coast of California. The overall error for this coastal error is shown in Table 5 and is about twice of what it is for the whole test area

Table 5 Comparison with 229,753 marine observations within 100 km from the coast. All values are in mGal.

	KMS2002	KMS99	GSFC	SS v 8
Mean	2.797	2.797	3.165	3.082
Std Dev	10.907	12.007	12.706	13.063
Minimum	-76.245	-77.346	-79.483	-87.651
Maximum	131.670	137.691	140.350	139.108

The overall error for this area is relatively large compared with other global comparisons like (i.e., Behrend et al., (1996), Tziavos et al. (1998), Andersen et al. (2001), Hwang et al. (2001)). The key result is the fact, that even with the improved ocean tide models the error close to the coast is about twice what it is for the whole test area. This demonstrates the importance of still trying to improve the global marine gravity field in shallow water regions in the future and investigate possible other means of improving the altimetric observations like re-tracking. Upon comparing altimetric gravity field anomalies with marine gravity observations it is important to remember that the marine gravity observations are not "true" gravity. A recent publication by Denker and Roland (this issue) highlights the error associated with the marine gravity observations themselves.

4 Summary

Important steps in the derivation of the KMS2002 global marine gravity field have been described in this paper. These improvements counts improved quality and coverage of the altimetric height observations, improved gravity field mapping in coastal regions, and improved modelling of the medium wavelengths geoid errors at 100-500 km. Extensive comparisons with marine gravity observations revealed, that the error in the altimetric derived gravity field close to the coast is still about twice what it is in the open ocean. This demonstrates the importance of still trying to improve the global marine gravity field in shallow water regions in the future.

Acknowledgment. The authors would like to thank R. Scharroo and colleagues for establishing the Radar Altimetry Data Service (RADS) at DEOS TU-Delft, and Anne L. Vest in preparing one of the figures.

Information. The KMS2002 gravity field is available from the authors on CD-ROM (oa@kms.dk) or from anonymous ftp at: ftp://ftp.kms.dk/pub/GRAVITY, www.research.dk/GRAVITY.

References

Andersen, O. B., and P. Knudsen, Global Marine Gravity Field from the ERS-1 and GEOSAT Geodetic Mission Altimetry, J. Geophys. Res., 103, 8129-8137, 1998.

Andersen, O. B., P. Knudsen and C.C. Tscherning, Investigation of methods for global gravity field recovery from dense ERS-1 geodetic mission altimetry. In: Global Gravity Field and its temporal variations. Ed Rapp, Cazenave & Nerem, IAG Symp 116, 218-226, Springer Verlag, Heidelberg, 1996.

Andersen, O. B., and Knudsen, P. The Role of Satellite Altimetry in Gravity Field Modelling in Coastal Areas, *Physics and Chemistry of the Earth (A)*, Vol. 25, No. 1, pp. 17-24., 2000

Andersen, O. B., P. Knudsen, S. Kenyon and R. Trimmer, Recent improvement in the KMS global marine gravity field, Boll Geofis. teor. ed Applic. Vol, 40, 3-4, pp 369-377, 1999.

Behrend, D. H. Denker and K. Schmidt, Digital gravity data sets for the Mediterranean Sea derived from available maps, Bull d.Information, 78, BGI, Touluse Cedex, France, 32 - 41,

1996.

Denker, H. and M. Roland, Compilation and evaluation of a consistent marine gravity data set surrounding Europe, This issue, 2004

Hwang, C., Hsu, H., and Jang, R. Global Mean Sea Surface and Marine Gravity Anomaly from Multi-satellite Altimetery: Applications of Deflection-geoid and Inverse Vening Meinesz Formulae, *Journal of Geod., in press.*, 2001

Lemoine, F. G., et al., The development of the NASA GSFC and DMA joint geopotential model, Proc. Int. Symposium on Gravity geoid and Marine Geodesy, Japan, Sept, 117,461-469, 1997.

Ray et al. A global Ocean Tide Model from TOPEX/POSEIDON altimetry: GOT99.2, NASA Tech Memo, 1999-2094478, GSFC, Greenbelt, 53pp, 1999.

Reigber, Ch., Balmino, G., Schwintzer, P., Biancale, R., Bode, A., Lemoine, J.-M., Koenig, R., Loyer, S., Neumayer, H., Marty, J.-C., Barthelmes, F., Perosanz, F., Zhu, S. Y.: A high quality global gravity field model from CHAMP GPS tracking data and Accelerometry (EIGEN-1S). Geophys.l Res. Lett. 29(14), 2002.

Scharroo, M. C. Naeije, and E. J. O. Schrama, Upgrading and harmonisation of ERS altimeter data: A problem solved? Geophysical Research Abstracts, 1(1), 217, 1999

Sandwell, D. T. and W. H. F. Smith, Marine gravity anomaly from Geosat and ERS-1 satellite altimetry, J. Geophys Res., 192, 10039-10054, 1997.

Shum, C. K., P. L. Woodworth, O. B. Andersen, G. Egbert, O. Francis, C. King, S. Klosko, C. Le Provost, X. Li, J. Molines, M. Parke, R. Ray, M. Schlax, D. Stammer, C. Tierney, P. Vincent, and C. Wunch, Accuracy assesment of recent ocean tide models, J. Geophys. Res., 102(C11), 25173-25194, 1997

Trimmer, R. G., Andersen, O., Driscoll, M.L., Wang, Y.M., Estimating Altimetry 5'x5' Mean Gravity Anomaly Regional Accuracies, In Vistas for Geodey in the New Millennium (ed). J. Adam and K.-P. Schwarz, Springer Verlag, IAG Symposia 125, 535-542, 2001

Tziavos, I. N., Forsberg, R., and Sideris, M.G Marine Gravity Field Modelling Using Ship-borne Gravity and Geodetic Misson Altimetry Data, *Geomatics Research Australasia,* Vol. 69, pp. 1-18, 1998

Wang, Y. M. GSFC00 mean sea surface, gravity anomaly, and vertical gravity gradient from satellite altimeter data., J. Geophys res., 106, C12, 31167-31174, 2001.

On the Determination of Marine Geoid Models by Least-Squares Collocation and Spectral Methods Using Heterogeneous Data

G.S. Vergos✉, I.N. Tziavos, V.D. Andritsanos
Department of Geodesy and Surveying, Aristotle University of Thessaloniki, University Box 440, 541 24, Thessaloniki, Greece, Fax: +30 31 0995948, email: vergos@topo.auth.gr

Abstract. In the frame of the EU-sponsored GAVDOS project the need of a new high-resolution and high-accuracy geoid model for the calibration of altimeters onboard satellites like JASON-1, ENVISAT and EURO-GLOSS and for sea level monitoring purposes has become apparent. That was mainly due to the fact that the already available models have been estimated using outdated datasets and fail to meet the wanted, cm-level, accuracy requirements. To determine the new geoid models multi-satellite (ERS1 and GEOSAT) altimetry and land and marine gravity data have been used. The EGM96 global geopotential model has been employed, while the effect of the bathymetry has been taken into account using recently developed local Digital Depth Models (DDMs). Several solutions have been estimated based on the different datasets used and the two main methodologies followed, i.e., the Fast Fourier Transform (FFT) based Input Output System Theory (IOST) and Least Squares Collocation (LSC). The accuracy of the new models was assessed through comparisons with TOPEX/POSEIDON (T/P) data and the GEOMED geoid solution for the area under study. Finally, the consistency between the estimated solutions has been determined by comparing the geoid height value they provide at the Gavdos Tide Gauge (TG) station on the isle of Gavdos. From the results it was found that the precision of the new geoid models is between ±0.9 and ±3.3 cm, their accuracy ranges between ±5 and ±10 cm and their consistency is at the ±0.5 – 6 cm level.

Keywords. Marine geoid modeling, least-squares collocation, IOST.

1 Introduction

The geoid serves as a reference surface for a wide number of Earth sciences and for various applications. Such are the calibration of altimeters onboard satellites, the determination and monitoring of the mean sea level and its variability and the deformation of the tectonic field, among others. For such applications to be successful a high-accuracy and high-resolution geoid model needs to be determined. Thus, in the frame of the EU-sponsored GAVDOS project, our group is participating with the aim of, among others, determining such a geoid model for the wider area of Gavdos and Crete. These two islands are located in the southern Aegean Sea, Greece, in an area with highly-variable gravity field features, due to the existence of the Hellenic Arc and the submergence of the African plate under the Euro-Asian one. Furthermore, the established calibration facility on the isle of Gavdos is unique since: a) it is under a crossing point of the T/P and JASON-1 tracks, b) it is adjacent to an ERS1/2 and ENVISAT pass, c) it is far from the mainland, and d) the sea circulation and the local tides in the area are relatively small. These facts make the isle of Gavdos ideal for calibrating altimetric satellites, sea level monitoring and tectonics studies, as long as a rigorous in terms of resolution and accuracy geoid model is determined. The first geoid model for the area was developed in the frame of the GEOMED project (1993) and further studies were conducted by Andritsanos et al. (2001b) and Vergos (2002).

In this paper some newly developed geoid models are presented and validated to asses their accuracy and precession. Thus, purely altimetric and gravimetric geoid models are determined, while two combined solutions are estimated as well using both traditional least-squares collocation and the FFT-based IOST method for the optimal combination of heterogeneous data. Furthermore, a rigorous processing methodology is followed, during which the effect of bathymetry and that of the quasi-stationary sea surface topography (QSST) are taken into account in the frame of the well-known remove-compute-restore method. Such effects have a significant influence on the accuracy of marine geoid models and need to be taken into account when cm-level accuracy is sought (Vergos and Sideris 2003a). The effect of the bathymetry needs to be removed for the residual geoid heights or gravity anomalies to be smooth before gridding or predic-

tion. According to Forsberg (1984), when high-quality depths are available, then the smoothing of the data can reach 50%. In marine geoid modeling a residual terrain model (RTM) reduction is used to account for the bathymetric effects (Forsberg 1984).

The effect of the QSST is important in processing altimetry data since the sea surface heights (SSHs) available from the satellites do not refer to the geoid but to the sea surface, thus their processing will determine a very good model of the mean sea surface but not the geoid itself. Thus, it is important to correct the altimetry SSHs due to the presence of the nearly-stationary part of the SST by simply removing its contribution.

The improvement that the newly estimated geoid models offer, compared to the previous ones developed for the Eastern Mediterranean Sea, is due to a) the use of more accurate satellite and terrestrial data, b) the higher resolution that can be achieved by combining different data sources, and c) the combination of heterogeneous data in the spectral domain with I/O algorithms.

2 Geoid Determination Methodology

Five geoid models, i.e., two purely altimetric, a gravimetric and two combined ones, have been estimated for the area under study. The development of the altimetric and gravimetric models as well as that of the two combined ones was based on the well-known remove-compute-restore method. The processing procedures for each case are well documented (see, e.g., Tziavos et al. 1998) so they will be briefly summarized here to place more emphasis on the numerical results.

2.1 Altimetric geoid modeling

The determination of a purely altimetric geoid model has been discussed in detail in Vergos and Sideris (2003a) and begins by using raw SSHs which have to be corrected for the various geophysical effects influencing the satellite measurements as well as the instrumental errors affecting the altimeters. This step results in the construction of Corrected SSHs (CSSHs) for each satellite mission. Since for some (e.g., GEOSAT), observations refer not only to oceanic but also land and shallow-water regions, a bathymetric mask has to be applied to remove the two latter. A depth limit can be set arbitrarily so measurements corresponding to depths smaller than that will be rejected. In the authors' opinion, the selection of this depth value is area-dependant and its smallness does not play a significant role, since if it is too small then there will remain erroneous observations in the data, most of which can be removed with a simple 3 rms test at a later step.

The so-derived SSHs refer to oceanic regions only and have to be reduced from the mean sea surface to the geoid by removing the contribution of the QSST, i.e., the nearly time independent part of the SST. Global SST models, which are based on altimetry and/or oceanographic data, or local solutions which employ in-situ measurements of salinity, pressure, temperature, etc, can be used to derive the QSST information. Once the SSHs are reduced to the geoid they can be regarded as geoid heights and are thus processed to give the final altimetric geoid model using the remove-compute-restore method. Thus, the contribution of a geopotential model is removed to derive reduced SSHs. However, these measurements still contain the influence of time varying oceanic effects and the radial orbit error. Such effects can be removed or reduced with crossover adjustment. Since the present study had a local character, a regional crossover adjustment scheme with one bias and one tilt parameter was applied (Rummel 1993). A local crossover adjustment can reduce not only the radial orbit error, mostly a bias in the data, but also some time-varying effects which are part of the influence of the sea state to the altimeter measurements.

To remove any remaining erroneous observations from the crossover adjusted residual SSHs, a simple 3 rms test was used, which was regarded to be sufficient for such purposes. The application of a 3 rms test assumes that the remaining errors in the data have a random nature, i.e., there are no biases left. This will be decided by examining the mean value of the reduced SSHs. If the mean value is small enough, e.g., below 0.10 m, the 3 rms test is performed and then the altimetry SSHs are RTM-reduced to derive the final residual SSHs. On the other hand, if the mean value is higher, then it means that some biases are still present in the data (even after the crossover adjustment) and should be removed prior to the 3 rms test. This can be done by first RTM-reducing the SSHs, since it is expected that the reduction will smooth the data (Forsberg 1984). This smoothness can be viewed by examining not only the mean value but the standard deviation (σ) of the reduced field too. The bathymetric effects should be accounted for with caution, since as shown by Vergos and Sideris (2003 a, b) if an inaccurate depth model is used then an error of about 2-6 cm is introduced in the final altimetric geoid. The residual SSHs available at this point were derived as

$$N_{res} = N_{obs} - N^{GM} - QSST - N^{RTM} \quad (1)$$

where N_{obs} denotes the observed SSHs, N^{GM} is the contribution of the geopotential model and N^{RTM} is the effect of the bathymetry. The so derived SSH_{res} represent the medium wavelengths of the geoid height signal and can be safely regarded as residual geoid heights N_{res}.

After that step, the final residual geoid height estimates can be derived by first gridding the data. This was performed for all solutions by LSC. The necessary variance and correlation length needed by the method were determined by computing the data empirical covariance function. The final step to determine the altimetric geoid model is to restore the contribution of the geopotential model and that of the bathymetry. The procedure described in this section is given schematically in Fig 1.

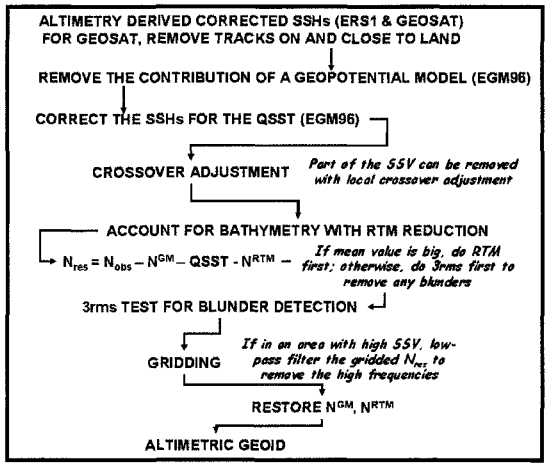


Fig. 1: Altimetric geoid modeling.

2.2 Gravimetric and combined geoid modeling

The determination of the gravimetric and combined geoid models was based on the processing scheme described in Sec. 2.1 and Fig. 1. The only exception was in the way the bathymetric information was handled in the former. In the gravimetric geoid determination the bathymetric reduction is applied in the remove step to smooth the residual gravity field, then the residual gravity anomalies are gridded and the contribution of the bathymetry is restored prior to the prediction of geoid heights. This is necessary since the bathymetry refers to masses below the geoid, which have to be restored before the use of Stokes's formula for geoid prediction. In the present study the prediction of the gravimetric geoid heights was carried out using the 1D-FFT method (Haagmans et al. 1993) and employing discrete spectra to evaluate Stokes' kernel function. No integration cap or kernel modification were used.

As far as the combined geoid modeling is concerned, two methods, i.e., the FFT-based IOST and LSC have been employed. The IOST solution was based on the equations given in Sideris (1996) and Andritsanos et al. (2001a) for the optimal combination of heterogeneous data. On the other hand the LSC model was determined using the well-known collocation formula given in Moritz (1989). In both cases the input signals were the two altimetric (ERS1 and GEOSAT) geoid models and the gravimetric one. Since there was no available information about the input errors, randomly distributed noise fields (white noise) were generated. The variance of each field was based on the σ of the differences of the respective geoid model with T/P SSHs.

3 Data for Geoid Model Estimation and Validation

The area under study is located in the Southern Aegean Sea centered on the island of Crete, Greece, $33^\circ \le \varphi \le 37^\circ$ and $21^\circ \le \lambda \le 29^\circ$. 174546 GEOSAT Geodetic Mission (GM), 105105 ERS1 GM and T/P altimetry data from the latest releases of their geophysical data records (GDRs) have been extracted for this area from the databases of NOAA (1997) and AVISO (1998), respectively. The gravimetric database comprised of a) marine gravity data available from the Institut für Erdmessung (Behrend et al. 1996) and 5′×5′ mean gravity anomalies from the GEOMED project databank and b) land gravity data provided by Lagios et al. (1996). These made up for a total number of 30437 gravity observations (see Figure 2).

The local depth models used to take into account the effect of the bathymetry were those developed by Vergos (2002) using satellite altimetry and shipborne depth soundings. Finally, the EGM96 global geopotential model, complete to degree and order 360, and the EGM96 DOT, complete to degree and order 20, were used to provide the long wavelength geoid information and the QSST respectively (Lemoine et al. 1998).

For the validation of the estimated geoid models, stacked T/P SSHs, known for their high accuracy, and the GEOMED geoid for the Mediterranean were used. The complete dataset of T/P SSHs from the 3rd year of its mission was extracted so as to have more reliable results. In all cases the differences between T/P or GEOMED and the estimated geoid were computed and minimized using a four-parameter transformation model:

$$N^v = N^i - b_o \cos\phi\cos\lambda - b_1 \cos\phi\sin\lambda - b_2 \sin\phi - b_3 \quad (2)$$

where the parameters b_o, b_1, b_2 and b_3 were calculated by a least squares technique, N^v denotes T/P SSHs or GEOMED and N^i denotes the altimetric *(i=a)*, gravimetric *(i=g)* or combined *(i=c)* geoid height depending on the solution under consideration.

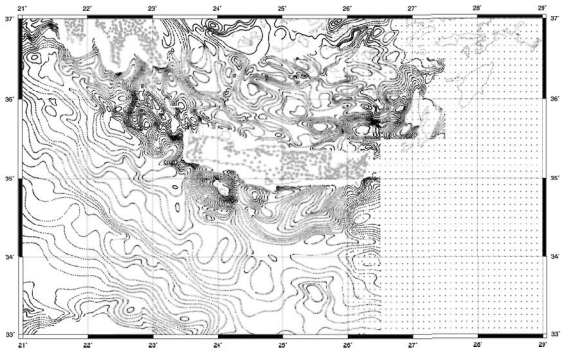


Fig. 2: Gravity data distribution (IfE: black, GEOMED: gray, Lagios: light gray circles).

4 Geoid Model Estimation

The development of the altimetric geoid models was based on the aforementioned ERS1 GM and GEOSAT GM data. Special emphasis was placed on crossover adjusting the data to reduce not only the orbital errors but sea variability effects as well. From that test it was concluded that even with the improvements in the satellite orbit determination, crossover adjustment is necessary since for both datasets it resulted in a reduction of both the mean and σ values by about 3-6 cm. Furthermore, it was found that by RTM-reducing the SSHs to account for the effect of the bathymetry the reduced field was much smoother (mean and σ reduction by about 4 cm) compared to the one prior to the reduction. This is a good indication that if a good DDM is available then it should be used to reduced altimetric SSHs prior to gridding or prediction (e.g., if the prediction of gravity anomalies is intended).

Following the processing scheme described in Sec. 2 the determination of all geoid models was performed on a 5′×5′ grid. It should be mentioned that to speed up the computations of the LSC-based combined solution, we decided to restrict it to a smaller inner area bounded between $33^\circ \le \varphi \le 37^\circ$ and $23^\circ \le \lambda \le 26^\circ$. This inner area was selected because it is centered on the isle of Gavdos which was our main focus. Table 1 summarizes the statistics of all five solutions. Note that the min and mean values in the LSC solution differ, because they refer to a smaller inner area.

Table 1 Statistics of the final altimetric, gravimetric and combined geoid solutions. Unit: [m].

	max	min	mean	σ
N^{GEOSAT}	38.10	0.67	16.59	±9.68
N^{ERS1}	37.98	0.73	16.68	±9.69
N^{gr}	38.04	2.09	16.99	±9.33
N^{IOST}	37.99	1.41	16.83	±9.51
N^{LSC}	33.20	5.83	20.17	±9.83

The differences between the gravimetric geoid solution and the altimetric ones is at the 15 cm level (1σ) and the highest values are concentrated in the eastern part of the area where the GEOMED data have been used. On the other hand the differences between the combined and altimetric models are smaller by about 8 cm something expected since the altimetry data have been used in the determination of the combined solutions. Figure 3 depicts the combined LSC solution.

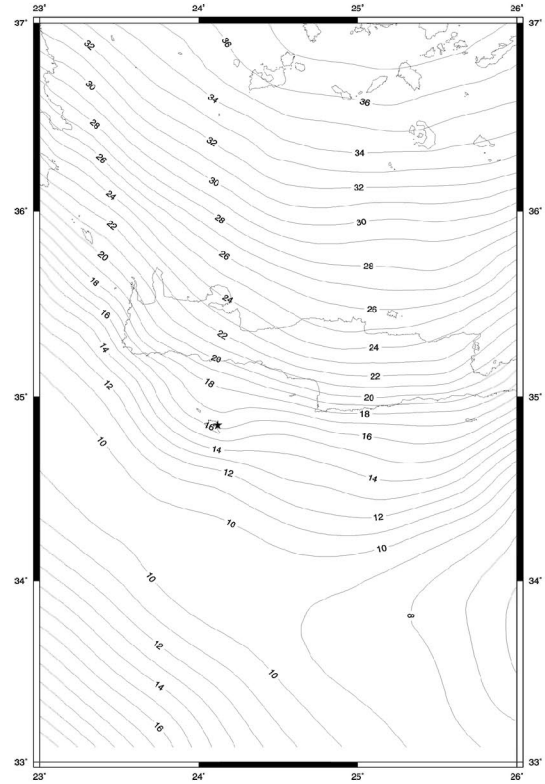


Fig. 3: The combined geoid model estimated with LSC (the black star shows the location of the Gavdos TG station).

5 Validation of the Estimated Geoid Models

The accuracy and precision of the models estimated were assessed by a) estimating the differences to T/P SSHs, b) estimating the differences to GEOMED geoid heights and c) comparing the geoid height estimate that each model gives for the Gavdos TG station. Table 2 presents the comparisons between T/P and the geoid models developed in this study. In all cases the differences were determined as $N^{T/P}$-N^{i} where i represents the geoid solution. The altimetric solutions agree with T/P at the ±10 cm level, which is better by about ±26 cm compared to the gravimetric geoid. The IOST and LSC solutions show an improved agreement with T/P by about 17 and 19 cm (1σ), compared to the gravimetric one, respectively. This leads to the conclusion that the combination of altimetry and gravity improves the accuracy of the gravimetric geoid. Additionally, it improves the altimetric geoid close to the coastline, where altimetry suffers from errors.

Table 2. Geoid height difference between T/P and the estimated models. Unit: [m].

	max	min	mean	σ
$N^{T/P} - N^{GEOSAT}$	0.34	-0.19	0.00	±0.11
$N^{T/P} - N^{ERS1}$	0.37	-0.23	0.00	±0.10
$N^{T/P} - N^{gr}$	1.27	-0.88	0.00	±0.36
$N^{T/P} - N^{IOST}$	0.59	-0.49	0.00	±0.19
$N^{T/P} - N^{LSC}$	0.40	-0.32	0.00	±0.17

Table 3 presents the geoid height differences between GEOMED and the estimated geoid models after the fit of a four-parameter transformation model. It is noticing that the range of the differences is in all cases close to about 2.7 m, which is quite high. Furthermore the σ ranges for all geoid models between ±42 and ±45 cm signaling that the GEOMED geoid model deviates from our solutions consistently. By plotting the differences with all geoid models it was found that they show some significant values throughout the area under study. This can be attributed to the different geopotential model, i.e., OSU91A, which was used in the development of the GEOMED geoid, in contrast to EGM96 that was employed in the present study.

Table 3. Geoid height difference between GEOMED and the estimated models. Unit: [m].

	max	min	mean	σ
$N^{GEOMED} - N^{GEOSAT}$	1.00	-1.74	0.00	±0.44
$N^{GEOMED} - N^{ERS1}$	0.99	-1.74	0.00	±0.43
$N^{GEOMED} - N^{gr}$	0.96	-1.82	0.00	±0.42
$N^{GEOMED} - N^{IOST}$	0.93	-1.72	0.00	±0.45
$N^{GEOMED} - N^{LSC}$	0.93	-1.75	0.00	±0.43

The final validation test was performed by estimating the geoid height value that each of the new models gives at the Gavdos TG station. For the combined LSC solution the geoid height was predicted by regarding the TG station as an additional estimation point, while in the other cases it was interpolated from the final solutions. The interpolation was performed in all cases using LSC and estimating the empirical covariance function for each individual solution to derive the necessary variance and correlation length. Table 4 presents the estimated geoid heights at the TG from the different solutions together with the precision in the prediction of each height. All estimates are very close and give a geoid height value at the Gavdos TG station of about 16.67 m. The differences between the predictions range from a minimum of 0.5 cm to a maximum of only 6 cm, while neglecting the gravimetric geoid height results in a maximum difference of only 2.8 cm. This signals that all solutions are consistent to each other and the methodology followed leads to precise results. The precision of the estimated height is ±3.33 cm in the worst case while it reaches the sub-cm level for the height from LSC. This, combined with the fact that the accuracy of the models is at the ±5-10 cm level, gives evidence of the rigorousness of the geoid models both in terms of the processing methodology and the accuracy and precision achieved.

Table 4. Estimated geoid height at the Gavdos TG station from the models developed. Unit: [m].

Model	N(m)	σ_N(cm)
N^{GEOSAT}	16.690	±2.47
N^{ERS1}	16.682	±2.05
N^{gr}	16.633	±2.82
N^{IOST}	16.662	±3.33
N^{LSC}	16.695	±0.92

6 Conclusions

With the aim of estimating an accurate and precise geoid model in support of the GAVDOS project, heterogeneous data and various methods have been used. This study aimed at providing the necessary methodological background so as to be ready to employ the additional gravity data that will become available to the project. The new data include marine, land and airborne gravity observations which will complement the existing gravity database. Thus, it was apparent that before implementing the entire database, different processing schemes had to be validated to choose the most rigorous one both in terms of accuracy and precision.

From the results of this study and the validation carried out, it became evident that when altimetry and shipborne gravity data are handled properly, i.e. corrected for all error sources, blunders removed, accurate geopotential and DOT models used, the data are corrected for the QSST signal, the bathymetry is taken into account using an accurate model, and the altimetry data are crossover adjusted, then, altimetric geoid modeling accurate to about ±5 cm is feasible. The combined solution is more accurate, compared to the gravimetric one, by about 17-19 cm, in terms of the σ of the differences with T/P, depending on whether LSC or IOST was used. It seems that LSC gives slightly better results (only by about 2 cm), but the IOST method is much faster. This last point becomes especially important taking into account that the complete gravimetric database by itself consists of approximately 100000 observations.

In terms of the accuracy of the models, the altimetric solutions performed much better compared to the gravimetric one, but that is probably due to the GEOMED data only. As shown by Vergos (2002) for the same area, when good quality marine gravity observations are used, then the gravimetric geoid is inferior to the altimetric ones by only 2-4 cm. Both combined geoid models (based on LSC and IOST) improved the gravimetric one in terms of the accuracy and the altimetric ones close to the coastline. The precision of the estimated geoid models ranged between ±0.9-3.3 cm, while they proved to be very consistent to each other, since the differences between the geoid heights that each one provided at the Gavdos TG were of the order of a few cm in the worst case.

Acknowledgements

This research was funded from the EU under contract GAVDOS: EVR1-CT-2001-40019 in the frame of the EESD-ESD-3 Fifth Framework Program.

References

Andritsanos VD, Sideris MG, Tziavos IN (2001a) A survey of gravity field modeling applications of the Input-Output System Theory (IOST). IGeS Bull 10: 1-17.

Andritsanos VD, Vergos GS, Tziavos IN, Pavlis EC, Mertikas SP (2001b) A High-Resolution Geoid for the Establishment of the GAVDOS Multi-Satellite Calibration Site. In: Sideris MG (ed) International Association of Geodesy Symposia, Vol 123, Gravity, Geoid and Geodynamics 2000. Springer, Berlin Heidelberg New York, pp 347-353.

AVISO (1998) User Handbook – Corrected Sea Surface Heights (CORSSHs). AVI-NT-011-311-CN, Edition 3.1.

Behrend D, Denker H, Schmidt K (1996) Digital gravity data sets for the Mediterranean Sea derived from available maps. BGI Bull d' Information 78: 31-39.

Forsberg R (1984) A study of terrain corrections, density anomalies and geophysical inversion methods in gravity field modeling. Report of the Dept. of geodetic Science and Surveying No 355. The Ohio State University, Columbus, Ohio.

GEOMED (1993) Mare Nostrum – GEOMED. Arabelos D, Tziavos IN (eds), Rep No. 4, Department of Geodesy and Surveying, Aristotle University of Thessaloniki, Thessaloniki, Greece.

Haagmans R, de Min E, van Gelderen M (1993) Fast evaluation of convolution integrals on the sphere using 1D FFT, and a comparison with existing methods for Stokes' integral. Manuscr Geod 18: 227-241.

Lagios E, Chailas S, Hipkin RG (1996) Newly compiled gravity and topographic databanks of Greece. Geophys J Int, 126: 287-190.

Lemoine FG, Kenyon SC, Factor JK, Trimmer RG, Pavlis NK, Chinn DS, Cox C, Klosko SM, Luthcke SB, Torrence MH, Wang YM, Williamson RG, Pavlis EC, Rapp RH, Olson TR (1998) The development of the join NASA GSFC and NIMA geopotential model EGM96, NASA Technical Paper, 1998 – 206861.

Moritz H (1989) Advanced Physical Geodesy, 2nd Edition. Herbert Wichmann, Karlsruhe.

National Oceanographic and Atmospheric Administration – NOAA (1997) The GEOSAT-GM Altimeter JGM-3 GDRs.

Rummel R (1993) Principle of Satellite Altimetry and Elimination of Radial Orbit Errors. In: Rummel R, Sansó F (eds) Satellite Altimetry for Geodesy and Oceanography, Lecture Notes in Earth Sciences No 50. Springer, Berlin Heidelberg New York, pp 189-241.

Sideris MG (1996) On the use of heterogeneous noisy data in spectral gravity field modeling methods. J Geod 70(8): 470-479.

Tziavos IN, Sideris MG, Forsberg R (1998) Combined satellite altimetry and shipborne gravimetry data processing. Mar Geod 21: 299-317.

Vergos GS (2002) Sea Surface Topography, Bathymetry and Marine Gravity Field Modelling. MSc Theses Dissertation, Department of Geomatics Engineering, University of Calgary, UCGE Reports 20157, Calgary, Alberta.

Vergos GS and Sideris MG (2003a) Estimation of High-Precision Marine Geoid Models Offshore Newfoundland, Eastern Canada. In: Tziavos (ed) 3rd Meeting of the International Gravity and Geoid Commission (IGGC), Gravity and Geoid 2002 - GG2002, pp 126-131.

Vergos GS and Sideris MG (2003b) Altimetry-Derived Marine Gravity Field Estimation Using Single- and Multi-Satellite Data. In: Tziavos (ed) 3rd Meeting of the International Gravity and Geoid Commission (IGGC), Gravity and Geoid 2002 - GG2002, pp 308-313.

Progress towards implementation and development of a New Zealand national vertical datum

M.J. Amos[1], W.E. Featherstone[2] and G.H. Blick[1]
1. Land Information New Zealand, Private Box 5501, Wellington, New Zealand; fax: +64 4 460 0112; email: mamos@linz.govt.nz
2. Western Australian Centre for Geodesy, Curtin University of Technology, GPO Box U1987, Perth, WA 6845, Australia

Abstract. New Zealand currently uses 13 disparate vertical datums, each connected to a separate tide gauge. In 1998, a new national datum, NZGD2000, was implemented based on GPS observations. This leads to a 3D geocentric datum. As further development of the spatial infrastructure in New Zealand, Land Information New Zealand approved the implementation of a new national vertical datum that is independent of local mean sea level. This new national vertical datum will be based on ellipsoidal heights, and the relationship between the separate existing vertical datums relative to the ellipsoid will be established using a high-precision regional gravimetric geoid model.

Phase one of this programme is the development of a regional geoid model for New Zealand. This paper will present the current status of development of the regional geoid model. Two geoids have been computed to determine the effectiveness of a 'gravity reconstruction technique' in New Zealand. The models computed are based on a combination of the EIGEN-2 satellite-only global geopotential model, which uses CHAMP dedicated satellite gravity data, and EGM96. Residual geoid undulations were computed from 40,000 land gravity observations and satellite altimeter-derived marine gravity anomalies. GPS and first-order spirit levelling data was used in conjunction with the geoid model to estimate offsets among the 13 vertical datums.

Keywords. Gravimetric geoid, vertical datum unification, New Zealand

1 Introduction

New Zealand does not currently have a single national vertical datum. Instead, 13 disparate local levelling datums are used. Each of these is based on local mean sea level (MSL) observed at different tide gauges at different times. The MSL determinations were generally over a short time period (less than two weeks in some cases). Despite some early evidence to the contrary, the datums were assumed to be stable and thus capable of being connected by precise levelling, so that they could eventually form a national vertical datum (e.g., Hannah, 2001).

Precise levelling is an accurate means of transferring heights between points, however the long period (~40 years) over which the New Zealand observations were made, means the network will most certainly be subject to uplift or subsidence due to a variety of processes. New Zealand is situated at the active boundary of the Australian and Pacific plates. This uplift and subsidence can be as much as 8.5 metres over ~40 years in localised areas, and ~10 mm/yr over larger areas (Walcott 1984).

Due to computational limitations at the time of calculation, each levelling network was adjusted independently. Even if a national adjustment was performed, New Zealand consists of two primary islands; so at least two vertical datums would result. The use of multiple local datums suited the height transfer technology of the time (i.e., precise levelling or vertical angles) and as most users only dealt with small spatial areas they were not concerned by the overlap (or lack thereof) of adjacent datums.

With the emergence of "modern" satellite based positioning technology (notably GPS) and the adoption of the New Zealand Geodetic Datum 2000 (see next section), both of which reference an ellipsoid as opposed to the local gravity vector, there is now a pressing need for a national height system that can incorporate both ellipsoidal and gravity-related heights.

This paper briefly presents the rationale for a new vertical geodetic datum in New Zealand, followed by recent geoid results, which advance upon those in Amos and Featherstone (2003a).

2 The Vertical Geodetic Datum

In 1998, the New Zealand Geodetic Datum 2000 (NZGD2000) was officially released as a replace-

ment for the existing New Zealand Geodetic Datum 1949 (NZGD49) (Grant and Blick, 1998). NZGD49 was a horizontal datum that was realised from over 35 years triangulation observations, coupled with astronomical observations for position and azimuth control.

In contrast to NZGD49, NZGD2000 is a 3D system that was realised using GPS observations and is defined in terms of the International Terrestrial Reference Frame 1996 (ITRF96) at the fixed epoch of 1 January 2000. To partially account for the tectonic dynamics of New Zealand, a horizontal deformation model forms an integral component of the datum. This model is used to transform observations backwards (and forwards) to 1 January 2000 so that consistent comparisons and adjustments can be made. It also assumes that there is zero vertical movement, which - as shown earlier - is clearly not the case. These and other limitations of the current datum, and some potential solutions, are discussed in Blick et al. (2003).

As stated earlier, NZGD2000 is a 3D datum that has its vertical component defined by the GRS80 ellipsoid. While this is a convenient mathematical reference surface, it is not easily conceptualised (or understood) by many users of the geodetic system. Users generally demand heights that are related to the gravity field, primarily to determine fluid flows and to be consistent with the existing geodetic infrastructure in New Zealand. It is therefore apparent that for a large number of applications (i.e., those that demand heights referenced to the local gravity vector) NZGD2000 is essentially a horizontal datum.

To establish a "usable" 3D datum, it is proposed to retain the ellipsoid as the "official" height system and to provide offsets between it and the 13 local datums. This will allow users to continue using the local datums, and will also support the use of technology, such as GPS, that references the ellipsoid. It will provide a nationally consistent height system for New Zealand. The obvious method of achieving this is through the use of a geoid model.

3 A Geoid Model for New Zealand

Unlike many other countries, New Zealand does not currently have a regional geoid model to support geodetic operations, including the transformation of GPS-derived heights to orthometric heights (e.g., Reilly, 1990). Global geopotential models are available to degree 360 (e.g., EGM96; Lemoine et al. 1998) that have an equivalent spatial resolution of ~50 km. This is insufficient for localised applications, as global geopotential models do not provide much information on the high-frequency variations in the geoid (i.e., the omission error).

Gilliland (1990) computed the first gravimetric co-geoid for New Zealand on a 0.25° grid using gravity data and the OSU81 (Rapp, 1981) global model to degree 180. Unfortunately this regional co-geoid model is no longer available for use. Nevertheless, advances in theory and data availability would now render this model outdated.

Therefore, there is a need to compute a new geoid model for New Zealand. Amos and Featherstone (2003a) computed a very preliminary co-geoid model, but it has since been found that some incorrectly mapped and thus pre-processed gravity data had been used. In addition, the above two co-geoid models omit the primary indirect effect term, which may be greater than ~0.5 m in magnitude at the summit of Aoraki/Mount Cook, the highest mountain in New Zealand (~3754 m above local MSL).

Probably the largest challenge to high-precision gravimetric geoid computation in New Zealand is its use of the 13 separate vertical geodetic datums based on local MSL. As stated earlier, each levelling network (cf. Gilliland 1987) is based on MSL observed at a different tide gauge, often over a very short time period. As well as not averaging out long-period tidal effects, these tide gauges are subject to sea surface topography (SST), which is notoriously difficult to quantify and model in the coastal zone (e.g., Hipkin 2000) or in harbours and estuaries where most of the tide gauges are located.

The effect of SST means that MSL measured at tide gauges departs from a single equipotential surface. This will create offsets between adjacent or overlapping vertical datums that have been based on such measurements (cf. Hipkin 2000). The broad-scale variation in SST estimated by satellite altimetry is 40 cm from the far North to far South of New Zealand (Hannah 2001).

Other oceanographic phenomena, such as storm surges or the outflow of fresh water (a number of gauges are located in river mouths) also act to bias the tide-gauge-measured MSL. Accordingly, the 13 vertical datums in New Zealand are not unified and may be offset from one another by more than 0.2 m (Pearse 1998).

These different vertical datums introduce two primary problems to practical geoid determination.

First, the regional gravity and terrain data used to compute the geoid model refer to different reference surfaces (i.e., MSL carried in-land by spirit levelling in each of the datums). This causes long- and medium-wavelength errors in the computed gravity anomalies (cf. Heck, 1990), which then propagate into the gravimetric geoid model. Secondly, a single transformation surface will not be suited for the direct transformation of GPS heights to these local vertical datums. Instead a different transformation surface for each datum will be required. Because the gravity data will be reduced onto different vertical datums it will propagate error into the surrounding areas, thus an iterative approach will be needed to compute the geoid model. Therefore, the New Zealand vertical datum will be based on a combination of ellipsoidal heights (in the 3D NZGD2000) and a precise regional geoid model. This geoid model will then allow the existing vertical datums to be unified (cf. Kumar and Burke 1998), but first it is necessary to contend with the above-mentioned practical and theoretical difficulties.

4 Preliminary Geoid Computations

Two new preliminary gravimetric geoids have been computed for the New Zealand region using a combined EIGEN2-EGM96 global geopotential model (see next paragraph), terrestrial gravity data and satellite derived marine gravity anomalies. The difference between the two models is in the way that the gravity data has been treated. One uses 'traditional' point free-air anomalies and the other uses mean free-air anomalies reconstructed using a DEM (described later).

Amos and Featherstone (2003b) tested fifteen global geopotential models (GGM) against gravity observations, GPS-levelling observations and deflections of the vertical in New Zealand and Australia. Included in the GGMs tested were two combined models that were constructed by replacing the low-order coefficients of EGM96 with those from EIGEN-2 (Reigber et al. 2002) and UCPH2002_02 (Howe et al. 2002) CHAMP-satellite-only models. They found a marginal improvement over EGM96 when the combined EIGEN2-EGM96 model was used. As such, this model is used as the reference GGM in this investigation.

The 2001 release of the Institute of Geological and Nuclear Sciences (GNS) national gravity database was used for the computation of this gravimetric geoid. This release consists of approximately 40,000 terrestrial and 1 million marine gravity observations.

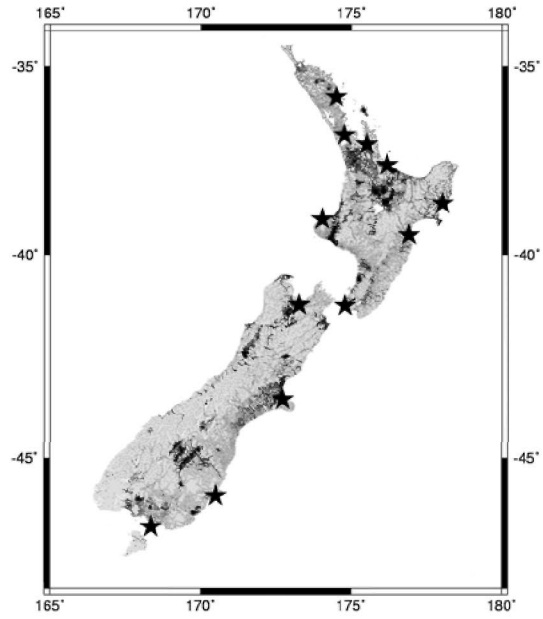


Figure 1. Tide gauge locations (indicated by stars) and terrestrial gravity coverage in New Zealand

The satellite altimetry grid of Hwang et al. (2002) was used to provide gravity information in the marine areas surrounding New Zealand. A comparison of the GNS marine data with the satellite altimeter data revealed that numerous errors exist as a result of no crossover analysis being performed. The crossover analysis is currently in progress, so for the purposes of this gravimetric geoid model the GNS marine data has been excluded and altimeter-derived anomalies used instead.

	Max	Min	Mean
Gravity	2023.6	107.9	492.164
DEM	2326.0	89.0	889.841

Table 1. Comparison of elevations from gravity observations and DEM in the Lewis Pass area of the South Island high country (units in metres)

The terrestrial gravity coverage (Figure 1) in New Zealand is, on average, spatially dense; however, the sampling method used is neither random nor regular. Due to the harsh topography in large parts of the country, the practicalities of collecting gravity data in the field mean that gravity is generally observed in the more accessible lowland re-

gions. Table 1 shows a comparison of the elevations of the gravity observations and elevations from a 1.8 arc-second (56 m) Geographx DEM in a one-degree square in the South Island high country.

It can be seen from Table 1 that the average height of the gravity observations is approximately half that of the DEM. This indicates that the locations of the gravity points are not representative of the topography in the area. This will result in an underestimate of the effect of topography in the geoid in this area.

Featherstone and Kirby (2000) describe a procedure where free-air anomalies are "reconstructed" from Bouguer anomalies by the application of a "reverse" Bouguer plate correction from the height of the DEM. This produces a grid of mean free-air anomalies at the same resolution as the DEM. For the purposes of this study, a 250-metre resolution Geographx DEM was averaged onto a 2-arc-minute grid. The abovementioned 56-metre DEM was not used for computational reasons; however, it will be implemented for future models.

The computation procedure employed for the two gravimetric geoids is described fully in Amos and Featherstone (2003c). In summary, two grids of free-air anomalies were compiled. The first arithmetically averaged the free-air anomalies extracted from the GNS database onto a two-arc-minute grid. The second averaged the Bouguer anomalies onto the same 2-arc-minute grid and then used the technique cited above to reconstruct the free-air values from a DEM on the same grid. These will be referred to as the "simple" and "reconstructed" grids respectively. Terrain corrections computed from Hammer charts to zone O were taken from the GNS database. No formal error estimate of these corrections has yet been made.

Both of the terrestrial gravity grids were then merged with the altimetry data to give two free-air anomaly grids over the computation area (160°E–10°W, 25°S–60°S). The EIGEN2-EGM GGM was then removed from each to give residual anomalies. The descriptive statistics of these anomalies are shown in Table 2. These differences are expected, as the reconstructed anomalies are more representative of the topography.

The residual anomaly grids were then subjected to a 1-D-FFT transformation with a deterministically modified Stokes kernel (Featherstone et al. (1998) with M=20 and ψo=1°) to evaluate the residual geoid. The resultant grids were then restored using the EIGEN2-EGM GGM and a correction for the primary indirect effect (Wicheincharoen 1982) was applied. The descriptive statistics of the various geoid contributions are shown in Table 3.

Grid	Max	Min	Mean	STD
Simple Mean	351.90	-292.80	1.75	34.36
Reconstructed Mean	555.19	-292.80	3.74	39.76
Residual Simple	352.53	-358.80	-0.15	13.41
Residual Reconstructed	539.18	-358.80	1.83	21.38

Table 2. Descriptive statistics of grided mean 2' by 2' gravity anomalies (units in mGal)

Grid	Max	Min	Mean	STD
EIGEN-EGM GGM geoid	54.051	-46.607	5.981	28.258
Residual Simple Undulations	2.208	-1.958	-0.019	0.226
Residual Reconstructed Undulations	15.355	-1.486	0.216	1.279
Indirect effect	0.000	-0.499	-0.001	0.007
Simple Geoid	54.131	-46.541	5.959	28.252
Reconstructed Geoid	54.169	-46.535	6.197	28.235

Table 3. Descriptive statistics of contributions to the New Zealand geoid (units in metres)

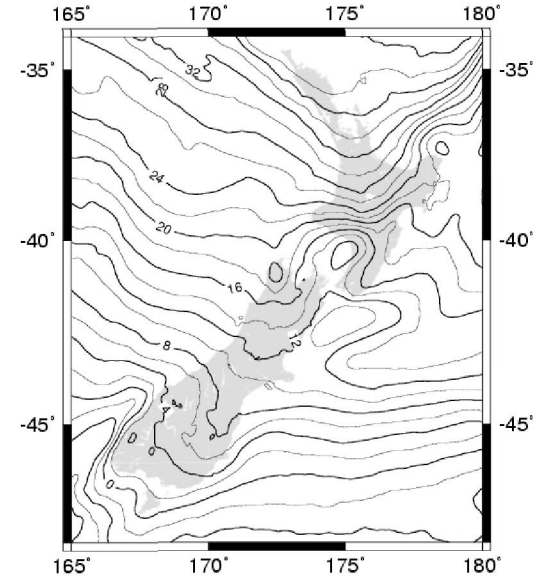


Figure 2. Gravimetric geoid of New Zealand using 'simple' gravity anomalies (contours in metres)

The resulting 'simple' and 'reconstructed' geoids are shown in Figures 2 and 3 respectively. The effect of the reconstruction can be seen very clearly in

Figure 3 where the Southern Alps (running north-east to southwest along the South Island) are now visible. This suggests that the effect of the topography is being treated more realistically.

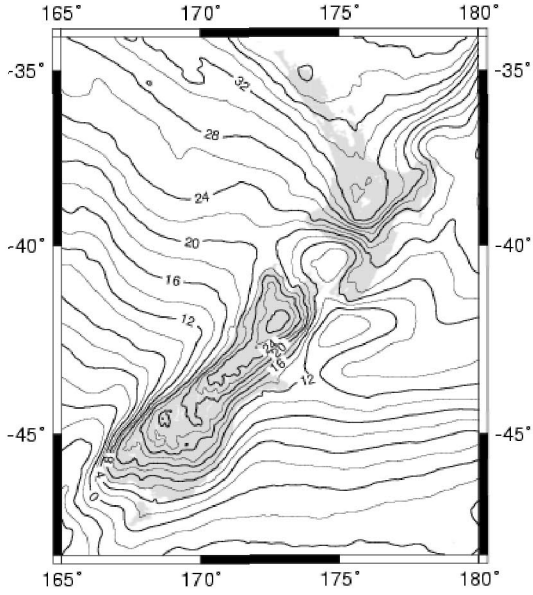


Figure 3. Gravimetric geoid of New Zealand using 'reconstructed' gravity anomalies (contours in metres)

The two geoids were then compared against a nation-wide set of 1055 GPS-levelling points. As mentioned earlier, the levelled heights are based on 13 different datums, which will bias the differences computed in Table 4. It is also pertinent to note that the effect of the solid Earth tide was ignored in the precise levelling reductions; this effect needs to be considered. Regardless of this, there is a significant improvement of the fit of the GPS-levelling to the 'reconstructed' geoid when compared to the 'simple' geoid. This suggests that the use of a higher resolution DEM in the reconstruction technique should be further investigated, as well as the role of topographical corrections.

Geoid	Max	Min	Mean	STD
EIGEN2-EGM	3.496	-1.376	-0.039	0.606
Simple	4.604	-0.480	0.436	0.754
Reconstructed	0.309	-1.712	-0.352	0.349

Table 4. Descriptive statistics of the comparison of geoid models with 1055 GPS-levelling points (units in metres)

The two geoids were then tested with GPS-levelling on a datum-by-datum basis. As well as to verify the new geoid models, this gives a *preliminary* indication of the offsets among these 13 vertical datums. The results of this testing with the reconstructed geoid are shown in Table 5. Though all the descriptive statistics are shown, only the mean differences should be interpreted as the preliminary offsets (cf. Featherstone 2000).

The values in Table 5 show broad agreement with observed differences between adjacent datums, which are in a restricted-access database held by Land Information New Zealand. However, the GPS-levelling points used in their derivation are often not representative of the entire datum. A program to acquire additional GPS positions at benchmarks is currently underway, the outcome of which will be a GPS-levelling point approximately every 7-10 km along levelling lines.

5 Conclusions

A new national vertical geodetic datum is proposed for New Zealand that will unify the 13 independent datums that are currently used. This datum will consist of a regional gravimetric geoid model that will provide the ability to transform heights from the local datums to the official ellipsoidal system. This will enable the retention of the disparate datums that are currently used to good effect and allow the use of technology such as GPS to obtain heights in terms of them.

Datum	Pts	Max	Min	**Mean**	STD
Auckland	84	-0.006	-0.612	**-0.373**	0.145
Bluff	91	0.280	-0.078	**0.050**	0.057
Dud-Bluff	170	0.309	-0.098	**0.114**	0.083
Dunedin	58	0.234	-0.778	**-0.288**	0.170
Gisborne	57	-0.508	-0.778	**-0.620**	0.080
Lyttelton	164	0.194	-1.712	**-0.556**	0.263
Moturiki	163	-0.087	-0.673	**-0.338**	0.146
Napier	26	-0.297	-0.551	**-0.438**	0.081
Nelson	46	-0.690	-1.185	**-1.021**	0.085
One Tree	34	0.001	-0.253	**-0.105**	0.065
Taranaki	57	-0.314	-0.726	**-0.535**	0.107
Tararu	13	-0.221	-0.775	**-0.530**	0.230
Wellington	67	-0.540	-0.964	**-0.816**	0.137

Table 5. Descriptive statistics of comparison of reconstructed geoid with GPS-levelling points on each vertical datum, datum offset in bold (units in metres)

The technique of 'reconstructing' free-air anomalies (Featherstone and Kirby 2000) using a 2-arc-minute DEM has shown significant improvements in accounting for topography. This is especially apparent in areas of harsh relief where gravity

observations are often made in the valleys and do not sample as thoroughly the more inaccessible surrounding mountains.

Of the two geoid models computed, the 'reconstructed' model has a significantly better fit to GPS-levelling than the simple model and other previously computed prototype models. The standard deviation of the differences is ~0.35 m for the 'reconstructed' model, which compares favourably to the ~0.75 m and 0.60 m from the 'simple' and GGM models, respectively. The computed offsets of each datum from the geoid are also better resolved. These estimates will be confirmed when a high-precision GPS survey at each of these points is completed later in 2003.

Acknowledgments. A Curtin University Postgraduate Scholarship and Land Information New Zealand fund this research. We would like to thank GNS, NIMA and C. Hwang for providing data, and the reviewers' for their comments.

References

Amos MJ, Featherstone WE (2003a) Progress towards a gravimetric geoid for New Zealand and a single national vertical datum. 3rd Meeting of the International Gravity and Geoid Commission, Gravity and Geoid 2002 - GG2002, Tziavos IN (ed), Thessaloniki, pp 395-400.

Amos MJ, Featherstone WE (2003b) Comparisons of global geopotential models with terrestrial gravity field data over New Zealand and Australia. Geom Res Aust *In press.*

Amos MJ, Featherstone WE (2003c) Preparations for a new gravimetric geoid model of New Zealand, and some preliminary results. NZ Surv 293: 9-20.

Blick GH, Beavan J, Crook C, Grant DB (2003 this issue) Practical limitations with the implementation of a semi-dynamic datum for New Zealand. Proc IUGG 2003, Sapporo, Japan.

Featherstone WE, Evans JD, Olliver JG (1998) A Meissl-modified Vaníček and Kleusberg kernel to reduce the truncation error in gravimetric geoid computations. J Geod 72: 154-160.

Featherstone WE (2000) Towards the unification of the Australian height datum between the mainland and Tasmania using GPS and AUSGeoid98. Geom Res Aust 73: 33-54.

Featherstone WE and Kirby JF (2000) The reduction of aliasing in gravity anomalies and geoid heights using digital terrain data. Geoph J Int 141: 204-212.

Gilliland JR (1987) A review of the levelling networks of New Zealand. NZ Surv 271: 7-15.

Gilliland JR (1990) A gravimetric geoid for the New Zealand region. NZ Surv 276: 591-595.

Grant DB, Blick GH (1998) A new geocentric datum for New Zealand. NZ Surv 288: 40-42.

Hannah J (2001) An assessment of New Zealand's height systems and options for a future height datum. Report prepared for Land Information New Zealand.

Heck B (1990) An evaluation of some systematic error sources affecting terrestrial gravity anomalies. Bull Geod 18: 227-241.

Hipkin RG (2000) Modelling the geoid and sea surface topography in coastal areas. Phys Chem Earth 25: 9-16.

Howe E, Stenseng L, Tscherning CC (2002) CHAMP gravity field model UCPH2002_02. Available from http://www.gfy.ku.dk/~stenseng/sagrada/poster.pdf

Hwang C, Hsu H-Y, Jang R-J (2002) Global mean sea surface and marine gravity anomaly from multi-satellite altimetry: applications of deflection-geoid and inverse Vening Meinesz formulae. J Geod 76(8): 407-418.

Kumar M, Burke KJ (1998) Realising a global vertical datum with the use of a geoid. In: Vermeer M, Adam J (eds) Rep 98:4, Finnish Geodetic Institute, Masala, pp 87-94.

Lemoine FG, Kenyon SC, Factor RG, Trimmer RG, Pavlis NK, Chinn DS, Cox CM, Klosko SM, Luthcke SB, Torrence MH, Wang YM, Williamson RG, Pavlis EC, Rapp RH, Olson TR (1998) The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) geopotential model EGM96. NASA/TP-1998-206861, NASA, Washington, 575 pp.

Pearse MB (1998) A modern geodetic reference system for New Zealand, UNISURV S-52, School of Geom Eng, UNSW, Sydney, Australia.

Rapp RH (1981) The Earth's gravity field to degree and order 180 using SEASAT altimeter data, terrestrial gravity data and other data: OSU81. Rep 332, Dept of Geod Sci and Surv, Ohio State Univ, Columbus.

Reigber C, Schwintzer P, Neumayer KH, Barthelmes F, Konig F, Forste R, Balmino G, Biancale R, Lemoine JM, Bruinsma S, Perosanz F, Fayard T (2002) The CHAMP-only EIGEN-2 Earth gravity model. Adv Space Res *Submitted.*

Reilly WI (1990) The geoid and the needs of the GPS user. NZ Surv 277: 35-41.

Walcott RI (1984) The kinematics of the plate boundary zone through New Zealand: a comparison of short- and long-term deformations. Geophys J R Astron Soc 79: 613-633.

Wichiencharoen C (1982) The indirect effect on the computation of geoidal undulations. Rep 336, Dept Geod Sci and Surv, Ohio State Univ, Columbus.

Prediction of Vertical Gravity Gradients Using Gravity and Elevation Data

Sz. Rózsa[1,2], Gy. Tóth[3]

[1] Geodetic Institute, University of Karlsruhe, Englerstr 7, D-76128
[2] On leave from the Department of Geodesy and Surveying and the Geodynamics and Physical Geodesy Research Group of the HAS, Budapest University of Technology and Economics
[3] Department of Geodesy and Surveying, Budapest University of Technology and Economics, Műegyetem rkp. 1-3, H-1111 Budapest and Physical Geodesy and Geodynamics Research Group of MTA-BME

Abstract. The vertical gravity gradients play an important role in the reduction of absolute gravity measurements and in the geoid determination, too. In order to enhance the precision of the gravity reductions and the geoid computations, the difference between the vertical gravity gradient of the real and the normal gravity fields should be taken into account.

In this paper the vertical gravity gradients are computed by prism integration using constant density for the crustal masses and a high-resolution (gridsize 10m) digital elevation model to represent the topography in the Sóskút test area. On the other hand, gravity gradients are also predicted from point gravity data using spectral techniques.

The calculated gradient values are compared to each other, and are validated using terrestrial measurements at 6 points in the Sóskút test area. The measurements of the vertical gravity gradients were carried out by the Loránd Eötvös Geophysical Institute and have an accuracy better than 80E, and are within the interval of 3189-3805E. The vertical gradient values calculated from the DEM range from 2926 to 3330E, and the values fitted the measured values with the standard deviation of ±244E, and the maximal residual of 547E.

The vertical gravity gradient has been investigated by Csapó - Papp (2000) in the study area, using the lithospheric model developed by Papp. Their results fitted to the observed values with the standard deviation of ±561E, with a maximal residual of –1298E.

1. Introduction

In 1999 and 2000 within the framework of a Research Programme of the Hungarian Academy of Sciences vertical gravity gradient measurements have been carried out at the six stations of the geodynamic network of the Department of Geodesy and Surveying in the Sóskút test area (Fig. 1). The network is located some 40 km west from Budapest in a steep valley underneath which it is suspected a fault is located.

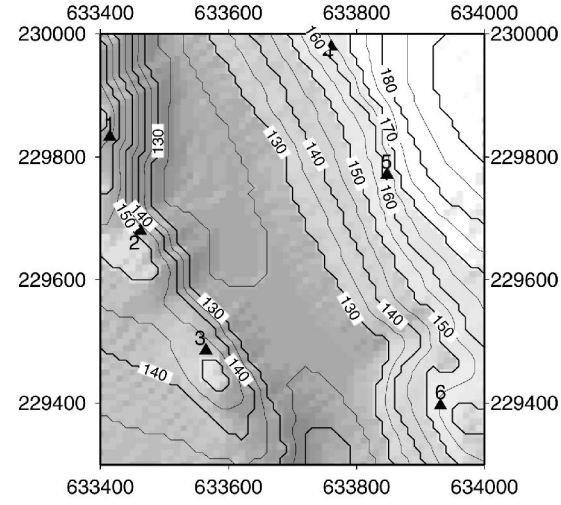


Fig. 1. The geodynamic network in Sóskút

The major goal of the research programme was to integrate various types of measurements in geodynamic analysis, such as spirit levelling, EDM measurements, traditional horizontal measurements, GPS, gravity measurements and a high-resolution geoid solution.

The gravity observations, and a high resolution DEM enable us to estimate the vertical gradients at the network stations. These values can be compared to the results of field observations.

2. Methodology

The computation of vertical gravity gradients have been split into three parts:

$$W_{zz} = U_{zz} + T_{zz}^{\Delta d} + T_{zz}^{top}. \quad (1)$$

The effect of masses below the geoid ($T_{zz}^{\Delta g}$) has been calculated from Faye-anomalies, in order not to duplicate the effect of topography. The effect of topographic masses (T_{zz}^{top}) has been computed using the DEM. U_{zz} is the free-air gravity gradient of the normal gravity field (3086E).

2.1. Effect of masses below the geoid

The effect of masses below the geoid has been computed according to the remove-restore technique. First the contribution of the EGM96 (Lemoine et al. 1998) has been removed from the the computed Faye-anomalies.

$$\Delta g_F^{res} = \Delta g_F - \Delta g_{GM} \quad (2)$$

Afterwards the contribution of these residual gravity anomalies (Δg_F^{res}), to the vertical gravity gradient has been calculated using the following formula:

$$T_{zz}^{res} = \frac{1}{4\pi R} \iint_{\sigma} \Delta g_F^{res} K_{zz}(r,\psi)\, d\sigma, \quad (3)$$

where

$$K_{zz}(r,\psi) = \sum_{l=2}^{\infty} \left(\frac{R}{r}\right)^{l+3} \frac{(l+1)(l+2)}{l-1} (2l+1) P_l(\cos\psi) \quad (4)$$

Eq. (4) can be expressed in a closed form for further computations:

$$K_{zz}(\sigma,x) = \sigma^3 \left(6\sigma^2 \frac{(x-\sigma)^2}{L^5} + \sigma \frac{11x-13\sigma}{L^3} + \frac{16}{L} - 18L + 18\sigma x \ln\frac{2}{M} - 45\sigma x + 2 \right), \quad (5)$$

where

$$\sigma = \frac{R}{r}, \quad (6)$$

$$x = \cos\psi, \quad (7)$$

$$L^2 = 1 + \sigma^2 - 2\sigma x, \quad (8)$$

$$M = 1 + L - \sigma x, \quad (9)$$

$$r = R + H, \quad (10)$$

R is the radius of the Earth, H is the elevation of the horizontal surface, where the results are needed. ψ is the spherical distance from the computation point to the moving point.

Due to the fact, that the kernel values are dependent only on the spherical distance and the elevation of the computational surface, the integral can be evaluated by FFT (Fast Fourier Transform) technique very effectively. However in this case the gravity anomalies must be available on a regular grid.

The normalized values of K_{zz} can be found in Fig.2. As it can be seen, the kernel function converges to zero quite quickly. This means that only a small neighborhood of the points must be taken into account. It is not important to collect gravity data for the computations beyond 300 to 400 m.

As mentioned before, the kernel values in (5) depend on the elevation, too. Therefore the computations have been carried out using two elevations bounding the elevation of the six network stations. Afterwards, the T_{zz}^{res} values have been interpolated using the T_{zz}^{res} values computed using these two elevations.

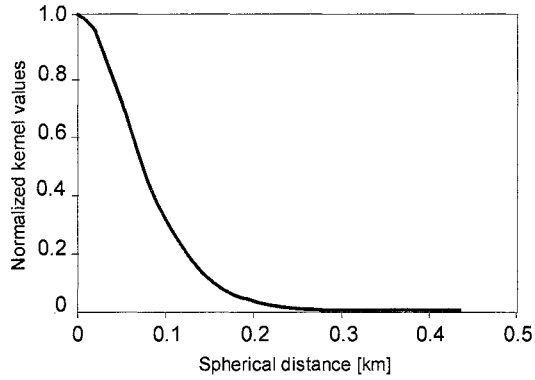


Figure 2. Normalized values of the kernel function for the computation of T_{zz} from gravity anomalies

After the calculation of T_{zz}^{res}, the contribution of the geopotential model can be calculated by the following formula:

$$T_{zz}^{GM}(P) = \frac{GM}{r^3}\sum_{l=2}^{l_{max}}\left(\frac{a}{r}\right)^l (l+1)(l+2)\sum_{m=-l}^{l} c_{lm} Y_{lm}(P) \tag{11}$$

where GM is the geocentric gravitational constant, c_{lm} are the coefficients of the geopotential model and Y_{lm} denote the spherical harmonics. The total effect of masses below the geoid is:

$$T_{zz}^{\Delta g} = T_{zz}^{res} + T_{zz}^{GM}. \tag{12}$$

2.2. Effect of topographic masses

The gravitational potential caused by a solid body can be expressed by the following formula:

$$V = k\iiint_v \frac{dm}{l} = k\iiint_v \frac{\rho}{l}\,dv, \tag{13}$$

where l is the distance between the mass element and the computational point, ρ is the density of the mass element and k is the gravitational constant.
The results of volume integration for the potential, and its derivatives for a right rectangular prism can be found in Nagy et al. (2000). After the cited article, the second derivative of the potential induced by a right rectangular prism with a constant density can be written as:

$$T_{zz}^{top} = G\rho\left|\left|\tan^{-1}\frac{xy}{zr}\right|_{x_1}^{x_2}\right|_{y_1}^{y_2}\Bigg|_{z_1}^{z_2}, \tag{14}$$

where x_1, x_2, y_1, y_2, z_1, z_2 are the bounding coordinates, and r is the Euclidean distance. Substituting all the limits the effect of the prism on the gravity gradient results in 8 terms.

One must handle Eq. 12. with care, because the second derivative of the potential does not exist in the whole space. However in these investigations, the vertical gravity gradients have been computed only in the exterior of the prisms, therefore no other considerations had to be made.

Due to the fact that the computations had to be performed only for six points, the simple numeric integration has been chosen to evaluate the effect of topography on the gravity gradient.

During the prediction of the vertical gravity gradients, the constant density of 2.67 g/cm^3 has been applied for all of the prisms represented by the DEM.

3. Data

3.1. DEM

The digital elevation model covers a 40 km by 40 km area around the geodynamic network, whose size is some 900 meters by 500 meters. The model has been created by the Mapping Agency of the Hungarian Armed Forces from topographic maps with the scale of 1:10000. The data is given on a regular grid on the Hungarian Unified National Projection System. The grid resolution is 10 meters.

The statistical properties of the DEM can be found in Table 1, while the DEM can be seen on Fig. 3.

Min	Max	Mean	Std. dev.
85	558	168.45	±71.96

Table 1. Statistical properties of the DEM [m]

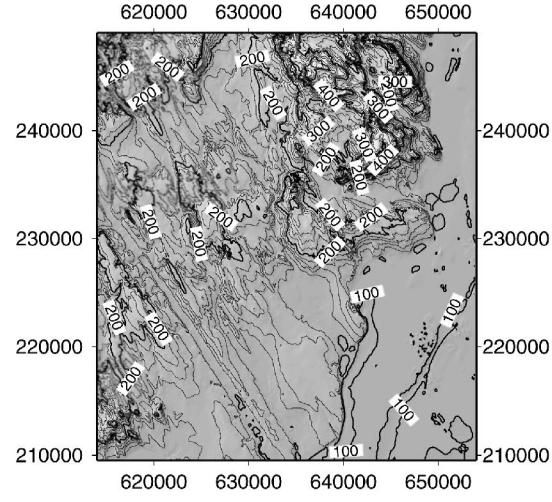


Fig. 3. The DEM of the study area

In Fig 1. one can see one of the main problems with the DEM. In reality the left hand side of the network (points 1-3) is situated close to the edge of an almost vertical cliff, while the DEM shows a height difference of 30 meters in 60(!) meters,

which gives a gradient of 26 instead of nearly 90 degrees. This error can be caused by the lack of detail of the topographic maps, and the interpolation technique used by the Mapping Agency.

In order to control the datum of the DEM, 151 point elevation values have been compared to the DEM elevations. The 151 values were the elevations of the gravity densification network in the central 20 km by 20 km square of the study area. The histogram of the discrepancies (Fig. 4) shows normal distribution with a small bias. The statistical properties of the discrepancies can be seen in Table 2. The bias of the elevations is +32 cm, and the standard deviation is ±2.32 m.

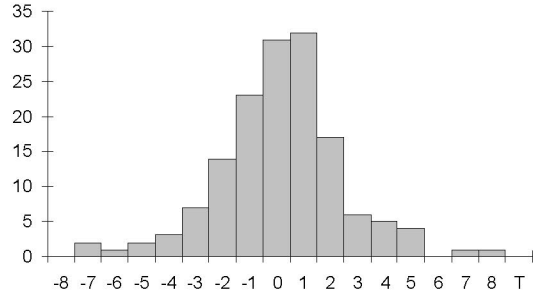


Fig. 4. Histogram of the discrepancies of point elevations and DEM elevations

Min	Max	Mean	Std. dev.
-7.30	7.90	-0.32	±2.32

Table 2. Statistical properties of the discrepancies from the comparison of DEM and point elevations [m]

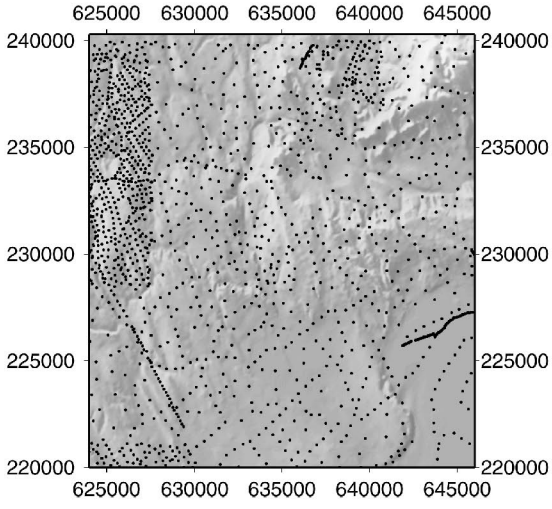


Fig. 5. The distribution of the gravity data

3.2. Gravity anomalies

Altogether 1610 point gravity data (Faye-anomalies) are available in the square of 20 km by 20 km around the network. Due to the fast convergence of the kernel function (Fig. 2.) only a subset of these must have been used for the calculations.

Unfortunately the distribution of the gravity measurements is not homogeneous (Fig. 5.), therefore only 41 point gravity data could be selected in the vicinity of the network. Of these, six observations have been taken at the stations of the network.

Using this 41 point gravity dataset, a rectangular grid has been created with the resolution of 33 meters. The statistical properties of the gridded Faye-anomalies can be found in Table 3.

Min	Max	Mean	Std. dev.
-12.57	7.68	-5.35	±4.91

Table 3. Statistical properties of the gridded Faye-anomalies [mgal]

3.3. Vertical Gravity Gradients

The vertical gravity gradients have been determined at all of the stations in the network by the Loránd Eötvös Geophysical Institute.

The measurements have been carried out using an LCR gravimeter set up at two different – but constant for each station – elevations on the vertical passing through the network stations. In order to achieve accurate results, the tidal effects, the drift of the gravimeter and the effect of atmospheric pressure has been taken into account.

The observations have been repeated twice at each station using two different gravimeters (except Station 1). The results can be found in Table 4. A detailed description of the applied method can be found in Csapó-Papp (2000).

Station	H[m]	No_{Sets}	No_{Gr}	W_{zz} [E]	M. err [E]
1	170	3	1	3599	20
2	157	2	2	3636	80
3	150	2	2	3805	79
4	169	2	2	3189	23
5	169	2	2	3203	74
6	163	2	2	3559	30
ΔW_{zz}=811E			Mean W_{zz}=3426 E		

Table 4. Vertical gravity gradients observed at the network stations (after Csapó-Papp, 2000)

4. Results

The effect of masses below the geoid has been computed for two elevations bounding the elevations of network stations, namely H=150 m and H=170 m. The statistical properties of the gridded T_{zz}^{GM} can be found in Table 5, while the spatial distribution of the values can be seen in Fig. 6.

Min	Max	Mean	Std. dev.
-222.80	151.43	-48.50	±71.89
-196.47	134.51	-46.19	±67.22

Table 5. Statistical properties of T_{zz}^{GM} for H=150m (first row) and H=170m (second row)

After interpolating the T_{zz}^{GM} values into the correct elevation of the stations, the predicted values of the vertical gravity gradients can be computed using Eq. 1. The various elements of the vertical gradients can be found in Table 6.

Station	$T_{zz}^{\Delta g}$ [E]	T_{zz}^{top} [E]	W_{zz}^{pred} [E]	W_{zz}^{obs} [E]	Diff. [E]
1	-34	0	3052	3599	547
2	-40	60	3106	3636	530
3	-54	298	3330	3805	475
4	-37	240	3289	3189	-100
5	-39	-121	2926	3203	277
6	-56	113	3143	3559	416
				Std. Dev = 244 E	

Table 6. Predicted vertical gravity gradients and their comparison with the observed values [E]

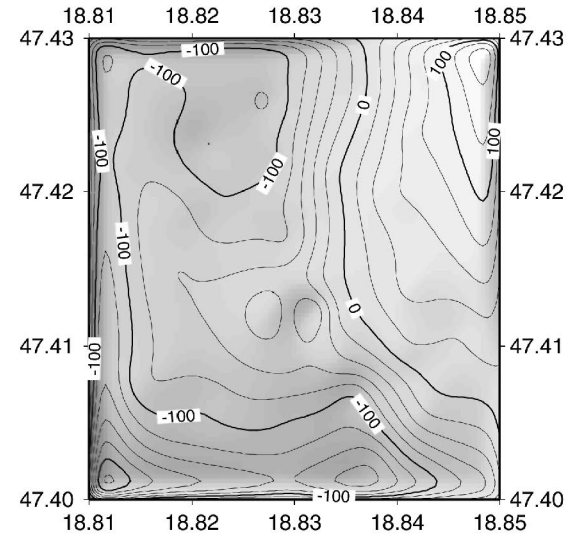


Fig. 6. The distribution of the T_{zz}^{GM} values

5. Conclusion and Outlook

Despite the situation of the network stations is not optimal for the computations, quite good results could be achieved. It should be noted, that although the overall standard deviation is ±244E, we can realize the standard deviation of ±37E for points 1-3 (hill on the left), and significantly worse standard deviation of ±267E for points 4-6 (hill on the right). One should also note the almost constant bias of the predicted values for points 1-3.

The prediction could be enhanced using a more precise DEM. The elevations obtained from the DEM are significantly lower than the exact point elevations (by almost 6 meters in terms of mean value). On the other hand, the topography is definitely not so steep in the DEM as in the reality. The hill on the left ends in an almost vertical wall on the eastern side, while the DEM shows only a slope with a gradient of 26 degrees, for reasons mentioned above.

On the other hand, lateral density variations could be also included in the computations in the near future.

Additional investigations are needed to explain the almost constant, and quite significant bias in the predicted vertical gravity gradients for the stations on the left hill. The reason for this bias could be the previously mentioned error in the DEM, or even the subsurface density variation. It is quite interesting to mention that in a precise geoid determination in the

test area (Tóth – Rózsa, 2000), the fit of the derived gravimetric geoid undulation to the height anomalies derived from GPS/Levelling measurements, showed significantly worse results for the stations 1-3 than for the station 4-6. The discrepancies had a mean value of 15.1 ± 1.3 mm and 7.6 ± 1.6 mm for the stations 1-3 and 4-6 respectively.

References

Csapó, G – Papp, G (2000): Measuring and Modelling of the vertical gradient of gravity – Hungarian examples (in Hungarian, Geomatikai Közlemények III. pp. 109-123.)

Lemoine, F.G. - Kenyon, S.C. - Factor, J.K. - Trimmer, R.G. - Pavlis, N.K. - Chinn, D.S. - Cox, C.M. - Klosko, S.M. - Luthcke, S.B. - Torrence, M.H. - Wang, Y.M. - Williamson, R.G. - Rapp, R.H. and Olson, T.R. (1998): The development of the joint NASA/GSFC and the National Imagery and Mapping Agency (NIMA) geopotential models, EGM96, NASA, TP-1998-206861, 575 pp, July 1998.

Nagy, D. – Papp, G. – Benedek, J. (2000): The Gravitational Potential and its Derivatives for the Prism (Journal of Geodesy, Vol 74., pp. 552-560.)

Tóth, Gy – Rózsa, Sz (2000): Local Gravimetric Geoid Modelling in the Sóskút Test Area (in Hungarian, Geomatikai Közlemények III., pp. 133-138.)

Construction of a Synthetic Earth Gravity Model by Forward Gravity Modelling

M. Kuhn, W.E. Featherstone
Western Australian Centre for Geodesy, Curtin University of Technology,
GPO Box U1987, Perth WA 6845, Australia, E-mail: kuhnm@vesta.curtin.edu.au, Fax: +61-8-9266-2703

Abstract. A synthetic Earth gravity model (SEGM) generates exact gravity field quantities and therefore is well suited to validate methods used in gravity field modelling. This paper describes the construction of a SEGM by forward gravity modelling using available data describing the Earth's mass distribution. In particular, the global 5-arc-min x 5-arc-min JGP95E digital elevation model (DEM), the most recent global compilation of crustal thickness and mass heterogeneities (CRUST2.0) and the S12WM13 mantle model of seismic velocity anomalies have been used. Results for the SEGM show in general a good agreement with the anomalous Earth's gravity field (geoid height). However, in some areas the differences with respect to the observed gravity field almost reach the magnitude of the geoid height itself. Possible explanations can be un-modelled mass anomalies situated in the upper and middle mantle.

Keywords. Synthetic Earth gravity model, forward modelling, Earth's mass-density distribution

1 Introduction

A synthetic [simulated] Earth gravity model (SEGM) can be constructed using either observation of the Earth's gravity field itself (effect model) or a reasonably realistic mass density distribution of the Earth's interior (source model) (e.g. Pail 1999). In the concept of effect models, the gravity field observations or derived gravity field quantities (e.g. gravity or geoid height) are taken as error-free model parameters. Effect models based on spherical harmonic expansions of the gravitational potential have been used by, e.g., Tziavos (1996), Novàk et al. (2001) and Featherstone (2002). The concept of source models, on the other hand, uses a reasonably realistic mass-density distribution in order to derive the gravity field directly through forward gravity field modelling. Most importantly, source models can describe the Earth's gravity field outside and inside the given mass distribution, whereas effect models generally only describe the external gravity field. Furthermore, a source model SEGM can be used for geophysical interpretations, provided that a realistic mass-density distribution has been applied. Therefore, the construction of a SEGM based on forward modelling, which is the focus of this paper, is of importance to both geodesy and geophysics.

Point-mass models provide simple source model SEGMs, which are very useful for generating the local and global gravity field. Such models have been developed by, e.g., Barthelmes and Dietrich (1991), Vermeer (1995) and Claessens et al. (2001). Global models based upon a spherical harmonic representation of Newton's integral have been developed by, e.g., Rummel et al. (1988) and Tsoulis (1999) using an idealised Earth's crust based on Airy's and Pratt's isostatic models. Kuhn and Featherstone (2003) applied successfully the global 2° x 2° CRUST 2.0 model (update of CRUST 5.1, Mooney et al. 1998) to construct a preliminary *crust-only* SEGM.

Using mass-density information only of the Earth's crust and implicitly assuming deeper masses to be homogeneous results in a SEGM with higher spectral power in the lower degrees in comparison to the observed gravity field (Kuhn and Featherstone 2003). This indicates the existence of additional mass anomalies located in the mantle, supposed to be responsible for the long-wavelength structure of the Earth's gravity field (e.g., Hager et al. 1985). Despite the generally higher amplitudes, some spatial pattern of the observed Earth's gravity field can be clearly identified in this preliminary *crust-only* SEGM, showing that they are mainly caused by crustal mass anomalies.

The models of Pail (1999), Haagmans (2000) and Claessens (2002) can be classified as hybrid source-effect models, as they combine both approaches. They take the long-wavelength structure from global gravity models (effect models) and only the short-wavelength part is derived from near surface mass-density information.

This paper presents the construction of a source model SEGM based on forward gravity field model-

ling using mass-density information of topography/bathymetry, crust and mantle. It is a logical continuation of the *crust-only* SEGM presented in Kuhn and Featherstone (2003), as it also includes mass heterogeneities of the mantle.

2 Methodology

The presented SEGM is based on forward gravity field modelling, thus the different gravity field parameters (here only the geoid height) of the SEGM are derived by evaluating Newton's volume integral for a given mass-density distribution. This can be done either in the space domain by numerical integration (*direct integration approach*) or in the frequency domain by expressing Newton's integral in a spherical harmonic expansion (*spherical harmonic approach*) (e.g. Kuhn and Featherstone 2003, Kuhn and Seitz 2003, this issue). The latter is well suited for global computations for the lower and medium spectral band (say degrees < 1000) of the gravity field, where the computation is normally less time consuming as in the direct integration approach. However, if higher spectral information is required (e.g., in local models), numerical integration is the only suitable approach, since the spherical harmonic approach suffers from numerical instabilities for degrees > ~2700 (Holmes and Featherstone 2002).

The numerical results for the SEGM presented in this paper are based on the spherical harmonic approach, which is preferred due to its efficiency in terms of computation time with respect to the direct integration approach. However, a final model should be based on the direct integration approach (especially for evaluation points on or near the Earth's surface) to include all spectral information available.

In the spherical harmonic approach, the gravitational potential $V(\Omega,r)$ caused by a given mass distribution at point $P(\Omega,r)$ located on or above the mean sphere with radius R is expressed by

$$V(\Omega,r)=\sum_{n=0}^{N_{\max}}\left(\frac{R}{r}\right)^{n+1}\sum_{m=-n}^{n}\bar{V}_{nm}\bar{Y}_{nm}(\Omega) \qquad (1)$$

with the fully normalised spherical harmonic coefficients $\bar{V}_{nm}$ and the surface spherical harmonics

$$\bar{Y}_{nm}(\Omega)=\begin{cases}\bar{P}_{nm}(\cos\theta)\cos m\lambda & m\geq 0\\ \bar{P}_{nlml}(\cos\theta)\sin|m|\lambda & m<0\end{cases} \qquad (2)$$

In Eq. (2) $\bar{P}_{nm}(\cos\theta)$ denotes the associated Legendre functions of the first kind for degree n and order m (e.g., Heiskanen and Moritz 1967). Here $\Omega=(\lambda,\theta)$ denotes the coordinate pair of spherical longitude λ and co-latitude (polar distance) θ and r stands for the radial distance measured from the geo-centre to the point $P(\Omega,r)$. The fully normalized spherical harmonic coefficients are given by an infinite sum of spherical harmonic coefficients $\bar{\kappa}_{p,nm}$ representing surface density functions (e.g. Ramillen 2002, Kuhn and Featherstone 2003)

$$\bar{V}_{nm}=\mu_n\left(\frac{R_L}{R}\right)^{n+2}\sum_{p=1}^{\infty}\frac{\Gamma(n+3)}{p!\Gamma(n+4-p)}\bar{\kappa}_{p,nm} \qquad (3)$$

with the factor

$$\mu_n=\frac{4\pi GR}{2n+1}. \qquad (4)$$

and the fully normalized spherical harmonic coefficients $\bar{\kappa}_{p,nm}$ of the surface density functions

$$\kappa_p(\Omega)=\rho(\Omega)\frac{H(\Omega)^p}{R_L{}^{p-1}},\qquad p\in N^+. \qquad (5)$$

In Eq. (5), the actual mass distribution used has been described by the laterally variable mass-density $\rho(\Omega)$ and the height ($H(\Omega) > 0$) and depth ($H(\Omega) < 0$) measured with respect to a mean sphere of radius R_L. Using a 3-D density function, the spherical harmonic approach can be either extended in radial direction (Ramillen 2002) or the density function can be transformed in a 2-D function by taking the integral mean along the height. The upward-continuation term $(R_L/R)^{n+2}$ in Eq. (3) shows that the effect of deeper mass distributions is attenuated for higher degrees, due to their greater distance from the geoid or the Earth's surface.

Interestingly, the spherical harmonic approach is somewhat insensitive to the vertical extension $H(\Omega)$, as it accounts only for the magnitude of the mass heterogeneity given by the surface density functions $\kappa_p(\Omega)$. Therefore, $H(\Omega)$ can simply be replaced by an equivalent rock height $H^{er}(\Omega)$ together with the constant reference density ρ_0 (e.g. Kuhn and Seitz 2003). However, this property can, in principle, be prevented by limiting the magnitude of $H(\Omega)$, which results in a subdivision of the given mass distribution into several concentrical layer structures accounted for by different upward continuation factors $(R_L/R)^{n+2}$.

The anomalous gravitational potential of the SEGM is given by the superposition of all effects caused by mass anomalies defined with respect to a reference model (e.g., Kuhn and Featherstone 2003). Choosing the centrifugal potential of the SEGM to be equal to the actual gravity field, the

anomalous (or disturbing) potential of the SEGM can be given by

$$T^{SEGM} = V^{\Delta m} \tag{6}$$

where $V^{\Delta m}$ indicates the combined effect on gravitational potential of all mass anomalies, given in the spectral domain by the spherical harmonic coefficients

$$\bar{T}_{nm}^{SEGM} = \bar{V}_{nm}^{\Delta m} = \sum_{i=1}^{I} \bar{V}_{nm}^{\Delta m_i} \tag{7}$$

where $\bar{V}_{nm}^{\Delta m_i}$ represents the spherical harmonic coefficients of the ith mass distribution $i = 1 \ldots I$.

In the spectral domain, other parameters of the anomalous gravity field can be derived from Meissl's spectral scheme (Rummel and van Gelderen 1995). Including the relation given by Eq. (3), an *extended* Meissl scheme for forward gravity field modelling can be defined (Kuhn and Featherstone 2003). Following Meissl's spectral scheme the effect on geoid height is given by (cf. Bruns)

$$\bar{N}_{nm}^{SEGM} = \frac{\bar{T}_{nm}^{SEGM}}{\gamma} \tag{8}$$

where γ denotes normal gravity on the ellipsoid.

In this paper, only the anomalous part $T^{SEGM} = V^{\Delta m}$ of the SEGM is studied. Therefore the results depend only on the introduced mass anomalies with respect to an assumed reference model.

3 Data: Synthetic Earth Mass Model

The construction of the SEGM is based on mass-density information of topography, bathymetry, crust and mantle. Each part of the mass-density information used will be briefly described including the adopted density reference model. Except the mantle all other sources have already been used in the construction of a preliminary *crust-only* SEGM given in Kuhn and Featherstone (2003).

Topography and Bathymetry

Topographic heights and bathymetric depths as well as information on oceanic and glacier ice and lake water are taken from the global 5-arc-min x 5-arc-min JGP95E digital elevation model (DEM; Lemoine et al. 1998, chap. 2). The different terrain types have been converted into equivalent rock heights (spherical approximation) using $\rho_0 = 2670$ kg/m^3 as reference density.

Crust

Crustal mass anomalies are taken from the global 2-deg x 2-deg CRUST 2.0 model, which (oddly) is an update of the CRUST 5.1 model (Mooney et al. 1998). In this model, the crust is defined by the depth and mass-density of five different layers (soft and hard sediments, upper, middle and lower crust). The bottom of the lowest layer defines the Mohorovičić discontinuity (Moho) used to model the topography of the crust/mantle transition zone.

Mantle

Density heterogeneities throughout the whole mantle (between 25 km and 2900 km depth) have been derived from the 3-D seismic shear-wave velocity heterogeneity model S12WM13 (Su et al. 1994). This model is expanded to spherical harmonic degree 12, to describe the horizontal variation, and to order 13 in Chebyshev polynomials to describe radial variations. The seismic velocity heterogeneity $\delta v/v_0$ has been expressed with respect to the preliminary reference Earth model (PREM, Dziewonski and Anderson 1981).

In this study $\delta v/v_0$ has been extracted every 25 km in depth between 25 km and 2875 km, expressed by their spherical harmonic coefficients $\delta v_{nm,i}$, which refer to the mean spheres of radii $R_i = R - i \cdot 25$ km [i = 1, ..., 115]. The set of all seismic velocity heterogeneities has been converted to mass-density heterogeneities (given by the surface density function in Eq. (5) for $p=1$ via

$$\bar{\kappa}_{1,nm,i} = \Delta r \frac{\rho_{0,i}}{\delta v / \delta \rho} \delta v_{nm,i} \tag{9}$$

where the factor $\delta v/\delta\rho$, which relates velocity changes to mass-density changes, has been set to 10, thus assumed constant throughout the mantle. The reference density $\rho_{0,i}$ for each depth is that of PREM and Δr indicates the vertical extension of each layer (here 25 km). Using Eq. (3), the resulting mass-density distribution given by a total of 115 layers of 25 km thickness have been converted into spherical harmonic coefficients of the gravitational potential by

$$\bar{V}_{nm,i} = G_{n,i} \mu_n \left(\frac{R_L}{R} \right)^{n+2} \bar{\kappa}_{1,nm,i} \,. \tag{10}$$

The dimensionless factors $G_{n,i}$ for each layer are Love numbers expressing the gravitational effect of a dynamic mantle (e.g. Kogan and McNutt 1993). In this study, a set of Love numbers, which have been assumed constant for each layer ($G_{n,i} = G_i$) has been derived by a simple fit (trial and error) of T^{SEGM} (degree 2 – 12) to EGM96 (degree 2-12).

Reference Density Model

The reference density model is assumed to be spherical with concentric layers of constant density. Within the crust of 25 km thickness, the reference crustal density is $\rho_c = 2861$ kg/m^3, which is the

mean density of CRUST 2.0. Beneath the reference crust the mass-density distribution of the PREM has been adopted. Finally, the mean density of masses below the core mantle boundary (CMB) has to be chosen in such a way that the reference model represents the remaining mass of the Earth.

4 Results: The SEGM geoid height

Numerical results given for the SEGM geoid height are based on the spherical harmonic approach [Eq (8)]. The *crust-only* SEGM described in Kuhn and Featherstone (2003) has been recomputed here using a slightly different mean crustal thickness of 25 km instead of 22 km.

The resulting geoid height of the *crust-only* SEGM illustrated in Fig. 1 shows several patterns of the observed geoid given by EGM96 (Lemoine et al. 1998) illustrated in Fig. 2. The spatial pattern is dominated by the effect of topography, bathymetry and the topography at the Moho. However, the amplitude range is in general much higher. This can be seen for example within the North Atlantic region, where the geoid height over Iceland shows nearly the same pattern as the observed geoid but the magnitude of over 150 m is more than two times larger than the observed 70 m. It shows that in this region the spatial pattern of the geoid is caused by the crustal mass distribution. However, the magnitude presumably depends also on deeper mass anomalies in the mantle. This is confirmed by the presence of a low-velocity zone in the upper mantle beneath the North Atlantic with its maximum beneath the Iceland plume (Foulger and Pearson 2001). Other spatial patterns, like the great low over Sri Lanka is completely missing in the *crust-only* SEGM. Therefore, deeper mass anomalies within the mantle have to be considered here in order to reduce the higher amplitudes and improve the spatial structure with respect to the observed geoid.

Fig. 3 shows the geoid height caused by the mass anomalies within the mantle derived from the model of shear-velocity heterogeneities S12WM13 (Su et al. 1993). Generally, it shows a similar pattern as the *crust-only* SEGM but with an opposite sign. Most of the spatial pattern in Fig. 3 is caused by mass anomalies in the first few hundred kilometres of the upper mantle. The effect of deeper mass-anomalies is smaller in magnitude and shows mainly long-wavelength structures corresponding to degree 2 and 3 (not shown here), which is consistent with the results presented in Hager et al. (1985). Combining the *crust-only* SEGM with the effects caused by the mantle mass anomalies results in a significant reduction of the amplitude range. This can be seen in Fig. 4, where the geoid height of the combined SEGM (called simply SEGM) using all data described in Section 3 is illustrated.

The resulting SEGM shows much smaller amplitudes with a maximum range of about ± 140 m, which is reasonably similar to that of the observed geoid. Most spatial patterns are now similar to the observed geoid, which makes the SEGM more realistic than that given by Kuhn and Featherstone (2003) and therefore suitable for studies within the [simulated] Earth's gravity field.

The inclusion of mantle information results in a reduction of spectral power of the *crust-only* SEGM in the first few degrees (Fig. 5). Especially for degrees 2 to 6, a significant reduction can be noted, where the spectral power becomes very similar to that of EGM96. Due to the limitation of the mantle model used (degree 12), the spectral power of the SEGM becomes identical to that of the *crust-only* SEGM beyond degree 12. Beyond degree 100, the spectral power of the SEGM shows an almost identical behaviour to that of EGM96. The remaining spectral band for degrees between 6 and 100 shows larger differences to EGM96. This spectral band can be associated with mass anomalies that have not been modelled by S12WM13. While the information for degrees greater than 100 is mainly dominated by topography, bathymetry and crust, the mantle information only contains information up to degree 12. Therefore, the differences between both models occur mainly in the spectral range of degrees 10 to 100 (Fig. 5). Following the relation in Bowin (1983, Eq. 9), these spectral range can be (loosely) associated with mass anomalies situated in a depth range of 1000 km to 60 km.

The differences between the SEGM and EGM96 are illustrated in Fig. 6. Here a special colour bar has been used indicating differences greater than ± 50 m (positive: white, negative: black). It can be seen that most differences remain below ± 50 m. However, some differences have nearly the same magnitude as the geoid heights themselves. A preliminary examination of these differences shows that most of them can be explained by geophysical structures within the mantle. Their detailed study, however lies beyond the focus of this paper.

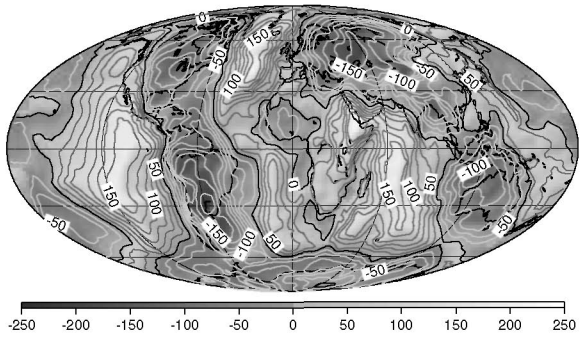


Fig. 1: Geoid height of SEGM crust only. Contour interval 50 m. [degree 2-360, min: -213 m / max: +216 m]

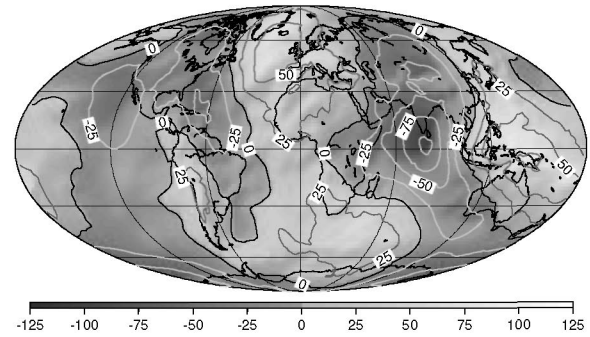


Fig. 2: Geoid height of EGM96. Contour interval 25 m. [degree 2-360, min: -105 m / max: +84 m]

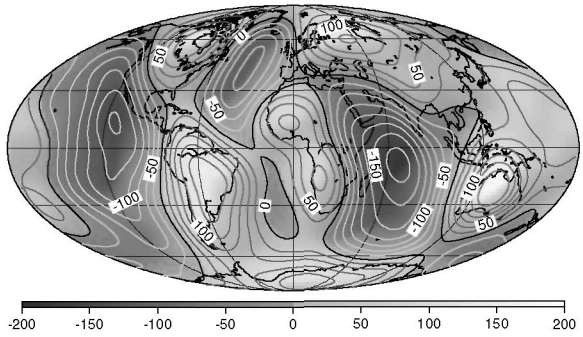


Fig. 3: Effect on geoid height from S12WM13. Contour interval 25 m. [degree 2-12, min: -185 m / max: +181 m]

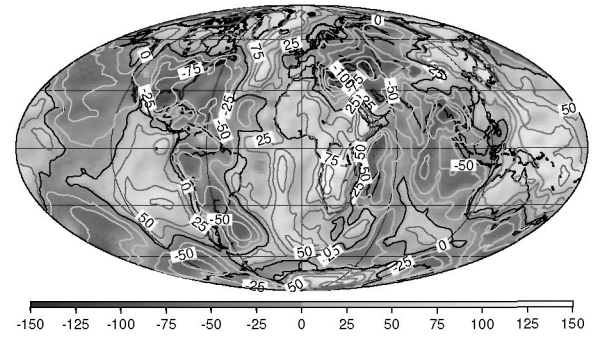


Fig. 4: Combined SEGM. Contour interval 25 m. [degree 2-360, min: -146 m / max: +140 m]

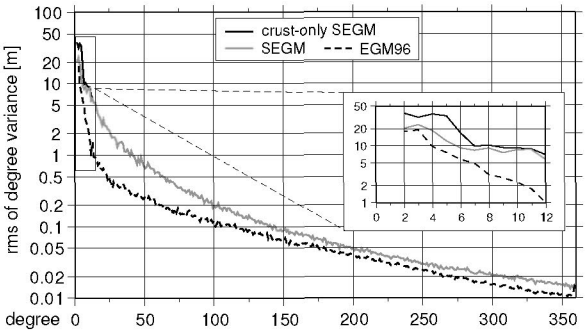


Fig. 5: Degree variances of *crust-only* SEGM, combined SEGM and EGM96.

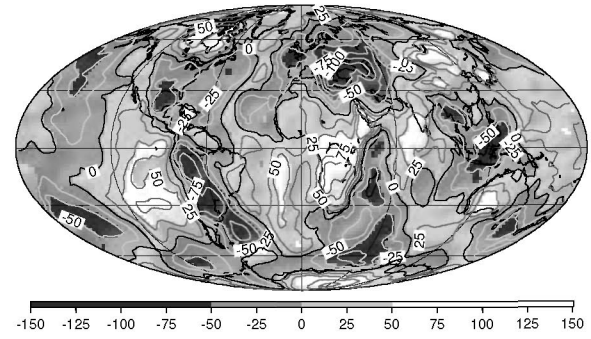


Fig. 6: Differences to EGM96. Contour interval 25 m. [degree 2-360, min: -147 m / max: +122 m]

5 Conclusions

It has been shown that the *crust-only* SEGM presented by Kuhn and Featherstone (2003) can be significantly improved by taking information on the mantle mass anomalies into account. The resulting SEGM is now realistic and can be used for simulation studies of the Earth gravity field. The remaining differences from the observed gravity field can be mostly attributed to a spectral range (degree 6 to 100), which is not or not very well presented by the data used. Furthermore, this spectral range can be (loosely) associated with mass anomalies situated in the upper and middle mantle (~60 km to ~1000

km). However, due to the non-uniqueness of the inverse problem there are also other (less likely) depth ranges possible (Bowin 1983).

Other sources that can explain the remaining differences can be found in the approximation of the factor $\partial v/\partial \rho$ (see Eq. 9) by a constant throughout the whole mantle as well as the Love numbers $G_{n,i}$ by a constant within each mantle layer. Lateral changes in viscosity for example can reverse the relation between seismic velocity and mass-density (e.g. Kogan and McNutt 1993).

In order to further improve the SEGM two approaches can be used: (1) Regional geophysical modelling or (2) assuming an (maybe un-realistic) mass-density distribution within the mantle, which inversely produces the pattern of the remaining differences (Fig. 6). The first way is clearly more difficult but it offers the possibility of a geophysical interpretation of the SEGM. The second way, instead, produces a nearly perfect match of the SEGM to the observed field (e.g. EGM96) but is geophysical meaning maybe reduced. However, the results of this study show that both ways produce a SEGM, which is realistic and can therefore be used for simulated studies of the Earth gravity field.

Acknowledgements This study was funded by the Australian Research Council through a grant A00001127. We thank Dr. K. Seitz (University of Karlsruhe, Germany) for software on spherical harmonic analysis.

References

Barthelmes L, Dietrich R (1991) Use of point masses on optimised positions for the approximation of the gravity field. In: Rapp RH, Sansó F (Eds) Determination of the Geoid, Springer, Berlin, pp 484-493.

Bowin (1983): Depth of principal mass anomalies contributing to the Earth's geoidal undulations and gravity anomalies. Marine geodesy **7**: 61-100.

Claessens SJ, Featherstone WE, Barthelmes F (2001) Experiences with point-mass gravity field modelling in the Perth region, Western Australia, Geom Res Aust **75**: 53-86.

Claessens SJ (2002) A synthetic Earth model analysis, implementation, validation and application, MSc thesis, Delft University of Technology, Delft, 75 pp.

Dziewonski AM, Anderson DL (1981) Preliminary reference Earth model. Phys Earth Planet Inter **25**: 297-356.

Featherstone WE (2002) Tests of two forms of Stokes's integral using a synthetic gravity field based on spherical harmonics. In: Grafarend EW, Krumm FW, Schwarze VS (Eds), Geodesy - The Challenge for the Third Millennium, Springer, Berlin, pp 163-171.

Foulger GR, Pearson DG (2001). Is Iceland underlain by a plume in the lower mantle? Seismology and helium isotopes. Geophys. J. Int. **145**: F1-F5.

Haagmans R (2000) A synthetic Earth for use in geodesy. J Geod **74**: 503-511.0

Hager BH, Clayton RW, Richards MA, Comer RP, Dziewonski AM (1985): Lower mantle heterogeneity, dynamic topography and the geoid. Nature **331**: 541-545.

Heiskanen WA, Moritz H (1967) Physical Geodesy, Freeman, San Francisco, 364pp.

Holmes SA, Featherstone WE (2002) A unified approach to the Clenshaw summation and the recursive computation of very high degree and order normalised associated Legendre functions, J Geod **76**: 279-299.

Kogan GM, McNutt MK (1993): Gravity field over Northern Eurasia and variations in the strength of the upper mantle. Science, **259**: 473-479.

Kuhn M, Featherstone WE (2003) On the optimal spatial resolution of crustal mass distributions for forward gravity field modelling. In Tziavos IN (ed) Gravity and Geoid 2002, Proceedings of the 3rd meeting of the International Gravity and Geoid Commission, Thessaloniki, Greece, 195-200.

Kuhn M, Seitz K (2003, this issue): Comparison of Newton's integral in the space and frequency domains. Paper presented at the General Assembly of the IUGG, June-July, 2003, Sapporo, Japan.

Lemoine FG and 14 other authors (1998) The development of the NASA GSFC and National Imaginary and Mapping Agency (NIMA) geopotential model EGM96. Rep. NASA/TP-1998-206861, National Aeronautics and Space Administration, Maryland, 575 pp.

Mooney WD, Laske G, Masters TG (1998) CRUST 5.1: a global crustal model at 5x5 degrees, J Geophys Res **103**: 727-747.

Novàk P, Vaníček P, Veronneau M, Holmes SA, Featherstone WE (2001) On the accuracy of modified Stokes's integration in high-frequency gravimetric geoid determination, J Geod **74**: 644-654.

Ramillen G (2002) Gravity/magnetic potential of uneven shell topography, J Geod **76**: 139-149.

Rummel R, Rapp RH, Sünkel H. (1988) Comparisons of global topographic/isostatic models to the Earth's observed gravity field, Rep 388, Dept Geod Sci and Surv, Ohio State Univ, Columbus.

Rummel R, van Gelderen M (1995) Meissl scheme – spectral characteristics of physical geodesy, manuscr geod **20**: 379-385.

Pail R (1999) Synthetic global gravity model for planetary bodies and applications in satellite gradiometry, PhD thesis, Technical University Graz, Austria, 138 pp.

Su W-J, Woodward RL, Dziewonski AM (1994): Degree 12 model of shear velocity heterogeneity in the mantle. J. Geophys. Res. **99(B4)**: 6945-6980.

Tsoulis D. (1999): Spherical harmonic computations with topographic/isostatic coefficients. Rep. No. 3 of the institute for astronomy and physical geodesy, Technical University of Munich, Germany, pp. 34.

Tziavos IN (1996) Comparison of spectral techniques for geoid computations over large regions, J Geod **70**: 357-373.

Vermeer M (1995) Mass point geopotential modelling using fast spectral techniques; historical overview, toolbox description, numerical experiment, manuscr geod **20**: 362-378.

First Results Towards An Isostatically Compensated Reference Earth Model

D. Tsoulis, B. Stary
Institute of Geodesy and Geoinformatics,
Berlin University of Technology, Sekr. H12, Straße des 17. Juni 135, D-10623 Berlin, Germany
Email: tsoulis@mca.bv.tu-berlin.de, Fax: ++49 30 314 21973

Abstract. Different isostatically compensated gravity models are assessed in view of a synthetic reference Earth gravity model. Among the models considered are a model produced by the analysis of the density and stratification information of the CRUST 2.0 global crustal database, and an isostatically compensated gravity model based on a simple spherical layer compensation mechanism. The deduced coefficients are used for direct computations of geoidal heights in a global scale. Appropriate filtering of the models relates the computed fields to known structures of the Earth's interior, like the topography of the core-mantle boundary (CMB).

Keywords. isostasy, gravity field recovery, reference Earth model

1 Introduction

Spherical harmonic analysis of global elevation models leads to a formal computation of potential coefficients describing the Earth's gravitational field. At the time being, global gravity field models, such as EGM96, correspond mainly to the long wavelengths of the geopotential field. Short wavelength information is still supplied regionally by surface gravity data, whereas the dedicated gravity field missions CHAMP, GRACE, GOCE and their successors aim at a highly precise coverage of long and medium wavelengths of the geopotential. GOCE, in particular, is expected to open a completely new range of spatial scales (order of 100 km) to the knowledge of the Earth's gravitational field.

A topographic/isostatic model, which results from the direct analysis of a global elevation model, should be sensitive to the intermediate wavelength band of the gravity field spectrum, as it incorporates a specific hypothesis for the isostatic balance between crust and upper mantle. In addition, it should contain long wavelength information as well, since its computation is based on the topographic data of a global elevation model. The balance between crust and mantle can be approxiamted by applying a certain compensation model for these density distributions. The application of the Airy/Heiskanen or Pratt/Hayford hypotheses provided in this respect satisfactory results (see e.g. Rummel et al. 1988, Tsoulis 2001), while for the evaluation especially of the high frequency content of topographic/isostatic models, one may incorporate surface gravity and satellite altimetry data as well (e.g. Pavlis and Rapp 1990).

Alternatively, one may use the information from recently published global crustal models, which provide data for density, compressional and shear wave velocities as well as thickness of a discrete number of crustal layers with a geological association (Kuhn and Featherstone 2003, Tsoulis 2003). The present paper attempts the assessment of different topographic/isostatic gravity models by means of a forward computation of geoidal undulations on a global scale. A filtered synthesis of some of these sets of potential harmonic coefficients reveals features that can be compared to the EGM96 geoid, or, in some cases, leads to potential fields that can be correlated to certain known geological features, such as the core-mantle boundary, also known as the topography of the D"-layer.

2 Isostatic potential

The topographic/isostatic potential is by definition the difference between the potential induced by an Earth which would behave as prescribed by a certain isostatic hypothesis, minus a reference potential induced by a simplified Earth model. The latter consists, in the conventional approach, of a crustal layer of uniform density ρ_{cr} and thickness D, superimposed on a denser underlayer of uniform density ρ_m (see also Pavlis and Rapp 1990). Both the isostatic hypothe-

sis and the adopted simplified Earth model may vary depending on the nature and structure of the available data. In the framework, for example, of the analysis of the crustal density and stratification information of the global database CRUST 2.0 (Bassin et al. 2000), one may apply both an Airy/Heiskanen and a radially distributed type of compensation mechanism, where the simplified Earth model is a two and a seven layer Earth model respectively (Tsoulis 2003). The definition of the topographic/isostatic potential T_{TI} remains thereby unaltered

$$T_{TI}(r,\vartheta,\lambda) = k \iiint_V \frac{\delta\rho(r',\vartheta',\lambda')}{l}\, du \quad (1)$$

where k is the universal gravitational constant, $\delta\rho(r',\vartheta',\lambda') = \rho_{hyp}(r',\vartheta',\lambda') - \rho_{model}(r',\vartheta',\lambda')$ the density disturbance at a source point, (r,ϑ,λ) and (r',ϑ',λ') the spherical coordinates of a computation and a source point respectively, l is the distance between computation and source point and finally V denotes the volume enclosing all density disturbances. Neglecting the presence of the atmospheric masses, T_{TI} is harmonic outside V, and its spherical harmonic series is convergent outside a sphere totally enclosing the disturbing masses. Under these assumptions, T_{TI} can be expressed as an infinite series in terms of fully normalized associated Legendre functions and unitless topographic/isostatic potential harmonic coefficients $C_{lm}^{T/I}$ and $S_{lm}^{T/I}$. A few remarks on the evaluation and assessment of the latter are presented in the following sections.

3 The CRUST 2.0 model

The use of global crustal data for geodynamic studies became recently very common among geoscientists, especially after the publication of the first relevant digital databases (Laske and Masters 1997). Based on the analysis of seismic refraction data gathered over almost fifty years and using a detailed compilation of ice and sediment thickness digital maps, Mooney et al. (1998) present a global model for the Earth's crust at a $5° \times 5°$ scale. The latest crustal model, CRUST 2.0, provides the same information (crustal density and stratification structure of the upper crust) with a $2° \times 2°$ resolution (Bassin et al. 2000). From the vast number of publications related to the use of these data, one should highlight here the direct computation of residual topography, gravity field and geoid in the frame of a global isostatic gravity model (Kaban et al. 1999), the derivation of a long-wavelength field for the Earth's dynamic surface topography (Pari 2001) and last but not least the construction of a synthetic Earth gravity model (Kuhn and Featherstone 2003).

Using the formal way of topographic/isostatic analysis, Tsoulis (2003) employed the CRUST 2.0 data to obtain the respective topographic/isostatic spherical harmonic potential coefficients up to degree and order 90. The analysis is performed in two ways: (a) using only the information for the Mohorovicic discontinuity depth, as given by the seventh layer of the CRUST 2.0 database and (b) using the overall stratification and density information for all seven layers provided by the same database.

Adopting an appropriate isostatic hypothesis (Airy/Heiskanen and a radially distributed compensation mechanism, respectively) the two approaches led to two sets of topographic/isostatic coefficients, complete to degree and order 90. The comparison of the two gravitational spectra with EGM96 and with the obtained spectrum based on the Airy/Heiskanen hypothesis with a constant compensation depth of 30 km, revealed the spectral content of the CRUST 2.0 information. Both CRUST 2.0 derived spectra appear to correspond to a relative under-compensated potential, with their degree variance curves being only slightly damped with respect to the spectrum of an uncompensated topography (for details see Tsoulis 2003).

The forward computation of global geoid undulations using the aforementioned coefficients affirmed their undercompensated character. The spherical harmonic synthesis of the overall coefficient set produces for both CRUST 2.0 derived models geoid undulation maps, having barely a physical meaning. Although the geographical distribution of the geoidal undulations looks similar to the respective EGM96 geoid, especially for the northern hemisphere, the values of the CRUST 2.0 derived geoid are so strongly scaled, that their geophysical or tectonical interpretation looks, at first glance, not possible. If, however, one defines certain wavelength bands of the CRUST 2.0 derived models, structures emerge that can be directly recognized as a residual geoid signal.

The upper panel of Figure 1 shows a filtered computation of CRUST 2.0 geoid undulations

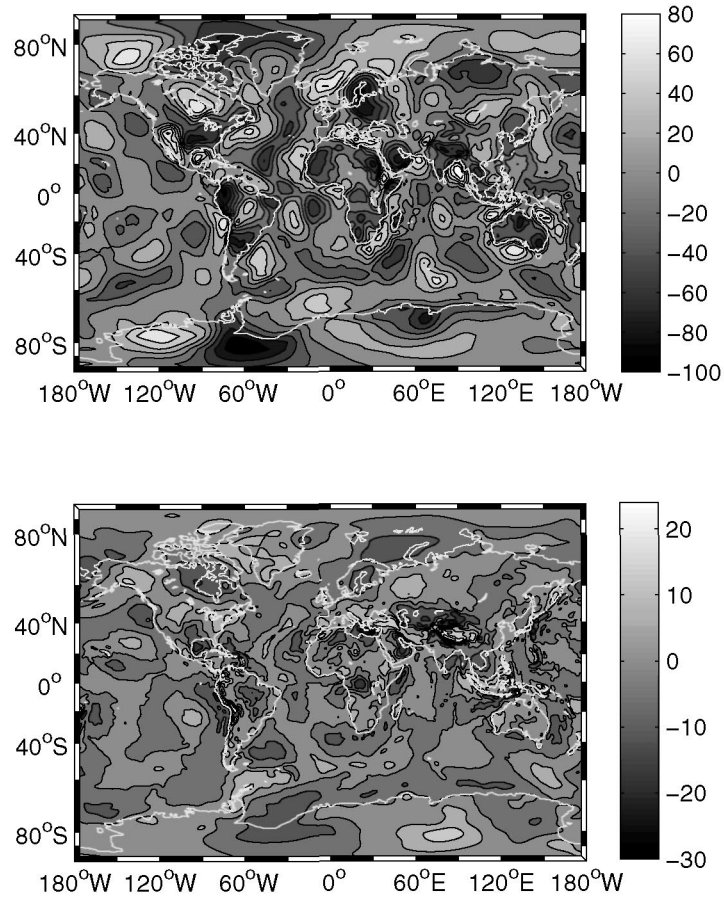


Fig. 1 Filtered CRUST 2.0 geoid undulations (7th layer only, $10 \leq l \leq 90$ and $0 \leq m \leq l$) applying an Airy/Heiskanen-like isostatic mechanism with $D = 22.97$ km (upper pannel) versus EGM96 geoid undulations for the same bandwidth (lower panel). Unit: m.

using the model emerging from the analysis of only the seventh crustal layer (global information for the Mohorovicic boundary). From the complete set of $C_{lm}^{T/I}$ and $S_{lm}^{T/I}$ coefficients only those were used for which $10 \leq l \leq 90$ and $0 \leq m \leq l$. For the formal computation of the coefficients an Airy/Heiskanen similar isostatic mechanism was applied (homogeneous crust of variable thickness lying above a homogeneous mantle) with a level of compensation equal to 22.97 km (Tsoulis 2003). The resulting residual geoid undulations vary between -100 and 80 meters. The lower panel of Figure 1 shows the EGM96 geoid undulations for the same bandwidth of coefficients. The magnitude of the EGM96 geoid values is considerably more moderate for the specific bandwidth. The CRUST 2.0 geoid values show a more apparent correlation with the continental part of the global topography. However, it is very interesting to observe, how the main geoid features (e.g. the transition from positive to negative values) are evaluated similarly from both models.

4 The spherical layer derived model

Tsoulis and Stary (2003) present an alternative and rather didactic approach for the computation of topographic/isostatic potential harmonic coefficients. They propose a simple hypothesis of spherical layers approximating constant density contrasts. Distinguishing between three basic surface topography data (continents, oceans and continents plus ice) an appropriate constant surface density is assigned to each elementary block of a global elevation model, according to

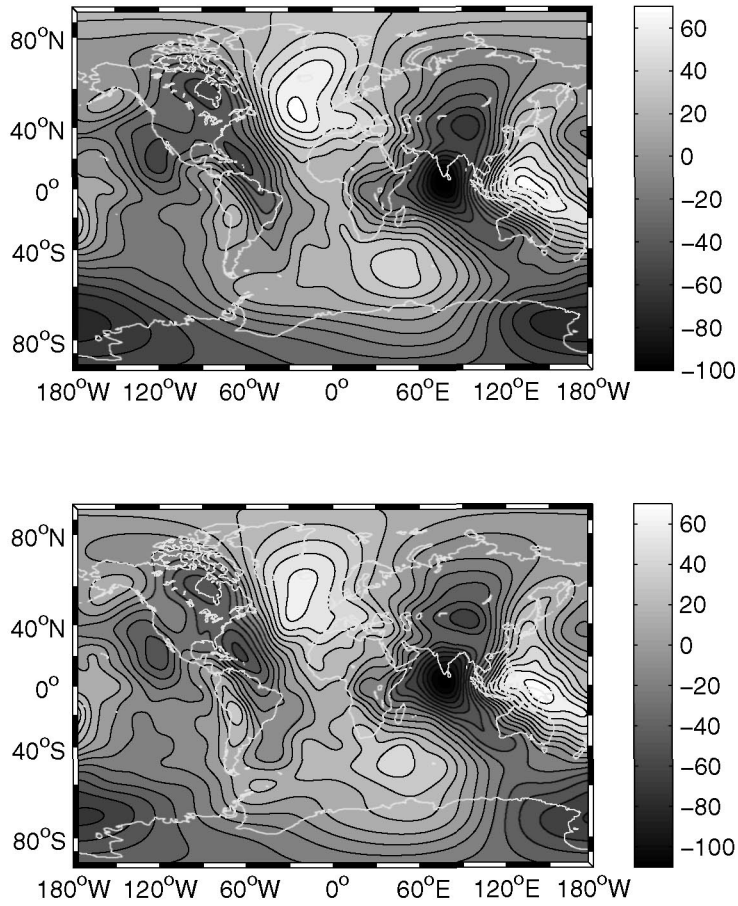


Fig. 2 Filtered geoid undulations ($0 < l < 21$ and $0 \leq m \leq l$) based on a spherical layer isostatic model with compensation depth $h_c = 50$ km (upper pannel) versus EGM96 geoid undulations for the same bandwidth (lower panel). Unit: m.

$\rho_1 = h \times \delta\rho$, where h denotes the discrete height or depth value as provided by the model, and $\delta\rho$ the respective constant density contrast, emerging from specific crude estimates of mean density values for the upper crust. The surface density which is evaluated in this way expresses the constant factor of proportionality of the mass deficiency or mass surplus spread over any part of the spherical layer to the area of that part of the surface. The isostatic mass balance principle, when applied to this layer model, dictates the existence of a negative mass equivalent ρ_2 at the compensation depth h_c, so that it will hold $\rho_2 = -\rho_1$. A similar concept has been applied from Kuhn and Featherstone (2003) in their definition of an equivalent mass layer.

The isostatic model emerging from this analysis was computed for three different compensation depths (30, 50 and 200 km). The power spectrum of the spherical layer compensation model for $h_c = 50$ km almost coincides with the spectrum of EGM96, for degrees l between 20 and 70, providing, in overall, a considerable amount of power decrease compared to the uncompensated topography spectrum (see Tsoulis and Stary 2003). Extracting exactly this frequency range ($21 \leq l \leq 70$ and $0 \leq m \leq l$) and computing the respective global residual geoid, one obtains a field which approaches zero for most of the globe, revealing amplitudes of up to $\pm$ 8 meters, mainly over regions of intensive tectonic activity. The upper panel of Figure 2 shows the residual geoid undulations computed by the bandwidth $0 < l < 21$ and $0 \leq m \leq l$ of the layer compensation model coefficients using $h_c = 50$ km. These undulations cannot be

readily explained through the occurrence of an isostatic mass compensation. The interesting resemblance of the structure of the undulations shown here with the topography of the D"-layer, presented in the following section, forms the assumption of assigning these large potential portions to long wavelength contributions stemming from structures in the Earth's interior, well below the crust-mantle boundary. Figure 2 also demonstrates the striking similarity of the layer model with EGM96 for the above mentioned coefficients bandwidth. The global mean difference between the two computed geoid fields amounts in absolute value 4.6 metres.

5 The D"-layer topography

The core-mantle boundary (CMB) is the most dramatic discontinuity in the Earth's internal structure, in terms of both the physical and chemical properties as well as the time scale of the processes that take place on either side of it (Morelli and Dziewonski 1987). By merging and inverting travel time residuals of waves reflected from and transmitted through the CMB, Morelli and Dziewonski (ibid.) published coefficients for the spherical harmonic expansion of the CMB-topography up to degree and order 4. The upper panel of Figure 3 presents the topography of the CMB, also known as the D"-layer, by invoking exactly these coefficients (ibid., Table 1). The lower panel of the same figure displays the geoid computed by the EGM96 coefficients up to degree and order 4. The correlation of the upper panels of Figures 2 and 3 is a first indication of connecting the long wavelength part of the computed layer gravity model of Tsoulis and Stary (2003) to the CMB structure.

The further and more in depth interpretation of the contribution of CMB topography to the observed gravity field signal demands the forward computation of the geoid signal implied by the CMB-defined density jumps. Forthcoming studies should clarify whether the respective potential contribution can be correlated with the long wavelength part of the observed gravity field (EGM96) or existing topographic/isostatic models. Critical issue in these investigations will be the definition of the lower limit of the disturbing masses, as CMB defines only their upper limit, and the choice of the modeling method for the computation of the respective disturbing potential for points at the Earth's surface. For example, a polyhedral modeling of these sources would be most suitable for an exact computation of their gravity field, while taking care at the same time of convergence effects of distant columns in a spherical approximation. This investigation can be applied to the CRUST 2.0 data as well. With the approximate geometric shape of each crustal layer given globally from the database, it is straightforward to translate these mass distributions to geoid signals. Thus, the contribution of each layer to the observed field can be analytically determined and the connection with the existing gravity models at different wavelengths directly realized.

6 Concluding remarks

The assessment of different topographic/isostatic geopotential models by computing geoid undulations in a global scale, revealed the importance of filtering the corresponding harmonic coefficients at different wavelength bands. The actual frequency content of global crustal databases, such as CRUST 2.0, could not be clearly estimated, with one of the reasons being surely the spectral limitation of the current model ($l_{max} = 90$). Crustal models of higher degrees, feasible only when denser crustal databases become available, are expected to reveal more spectral characteristics of the upper crust, increasing their applicability for geodesists and geophysicists. The spherical layer approach, on the other hand, provided a model, which is able to explain the medium wavelengths of the geopotential. For the short wavelength part of the spectrum, the interpretation should be found in contributions coming from heterogeneities in the mantle's structure, or at the CMB. An appropriate filtering of the currently available models at those bandwidths, where they are, or can be geophysically or tectonically interpreted, should yield, in time, an isostatically compensated reference Earth model.

References

Bassin C, Laske G, Masters TG (2000) The currrent limits of resolution for surface wave tomography in North America, EOS Trans AGU 81: F897

Kaban MK, Schwintzer P, Tikhotsky SA (1999) A global isostatic gravity model of the Earth, Geophys J Int 136: 519–536

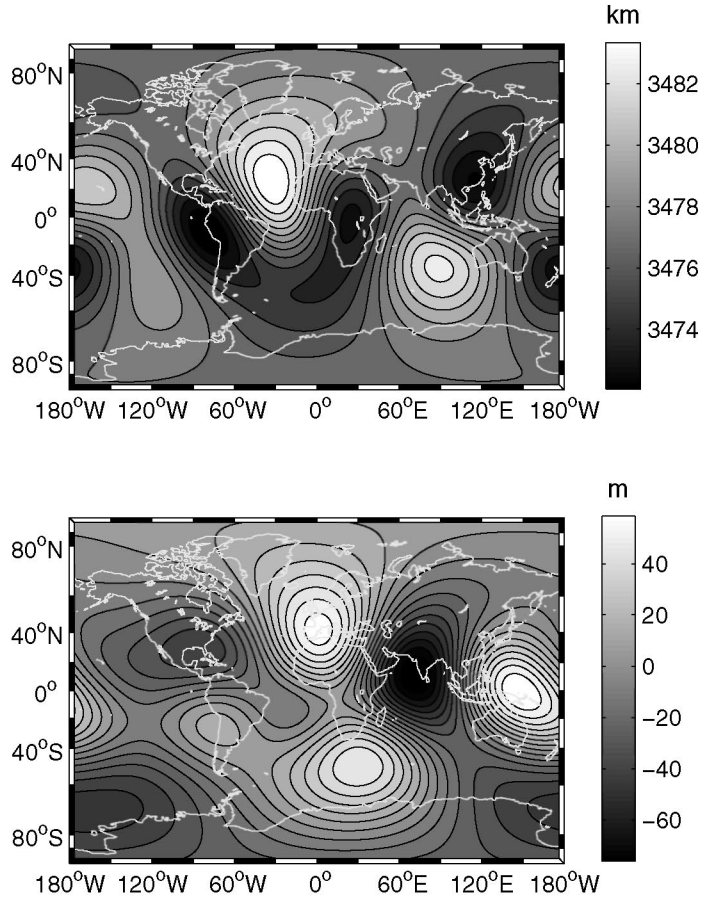


Fig. 3 Topography of core-mantle boundary using the coefficients up to degree and order 4 published by Morelli and Dziewonski (1987) (upper pannel; Unit: km) versus EGM96 geoid undulations using the coefficients $0 < l \leq 4$ and $0 \leq m \leq l$ (lower panel; Unit: m).

Kuhn M, Featherstone WE (2003) On the construction of a Synthetic Earth Gravity Model (SEGM), In: Gravity and Geoid 2002 (Ed. IN Tziavos), Proc. 3rd Meeting of the International Gravity and Geoid Commission, Thessaloniki, 26-30 August 2002, pp. 189–194

Laske G, Masters TG (1997) A Global Digital Map of Sediment Thickness, EOS Trans AGU 78: F483

Morelli A, Dziewonski AM (1987) Topography of the core-mantle boundary and lateral homogeneity of the liquid core, Nature 325: 678–683

Pari G (2001) Crust 5.1-based inference of the Earth's dynamic surface topography: geodynamic implications, Geophys J Int 144: 501–516

Mooney WD, Laske G, Masters TG (1998) CRUST 5.1: A global crustal model at 5° × 5°, J Geophys Res 103: 727–747

Pavlis NK, Rapp RH (1990) The development of an isostatic gravitational model to degree 360 and its use in global gravity modelling, Geophys J Int 100: 369–378

Rummel R, Rapp RH, Sünkel H, Tscherning CC (1988) Comparisons of global topographic/isostatic models to the Earth's observed gravity field, Rep 388, Department of Geodetic Science and Surveying, The Ohio State University, Columbus

Tsoulis D (2001) A comparison between the Airy/Heiskanen and the Pratt/Hayford isostatic models for the computation of potential harmonic coefficients. J Geod 74: 637–643

Tsoulis D (2003) Spherical harmonic analysis of the CRUST 2.0 global crustal model. J Geod, in press

Tsoulis D, Stary B (2003) An isostatic compensated gravity model using spherical layer distributions. J Geod, in review

Towards the Estimation of a Multi-Resolution Representation of the Gravity Field Based on Spherical Wavelets

M. Schmidt[1,2], O. Fabert[1] and C.K. Shum[2]

[1] Deutsches Geodätisches Forschungsinstitut (DGFI), Marstallplatz 8, D-80539 München, Germany, email: schmidt@dgfi.badw.de

[2] Laboratory for Space Geodesy and Remote Sensing Research, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210, U.S.A.

Abstract. Usually the gravity field of the Earth is modeled by means of a series expansion in terms of spherical harmonics. However, the computation of the series coefficients requires preferably homogeneous distributed global data sets. Since wavelet functions localize both in the spatial and in the frequency domain, regional and local structures may be modeled by means of a spherical wavelet expansion. Wavelet-based techniques allow the decomposition of a given data set into frequency-dependent detail signals. This paper deals with two methods to achieve a multi-resolution representation, namely a wavelet-only solution and a combined approach. The latter consists of a spherical harmonic expansion for the low-frequency part and a spherical wavelet expansion for the remaining medium- and high-frequency parts of the gravity field. Parameter estimation procedures are presented to determine the unknown series coefficients of both approaches.

Keywords. Spherical harmonics, spherical wavelets, multi-resolution analysis, parameter estimation methods.

1 Introduction

The determination and the representation of the gravity field of the Earth are some of the most important topics of physical geodesy. In the last years several approaches were pursued to generate a *multi-resolution representation* (MRR) of the geopotential; see e.g. Freeden (1999), Kusche (2002) and Beylkin and Cramer (2002).

The computation of the coefficients of a series expansion for the geopotential in terms of spherical harmonics requires preferably homogeneous distributed global data sets. Since a wavelet function is characterized by its ability to localize both in the spatial and in the frequency domain, regional or even local structures can be modeled by means of an appropriate wavelet expansion. Applying the wavelet transform, a given data set is decomposed into a certain number of frequency-dependent detail signals, i.e. an MRR is performed.

In this context there are generally two appropriate methods to model the geopotential, namely

1. the *wavelet-only approach*, i.e. the entire frequency spectrum is described by means of a series expansion in terms of wavelets and
2. the *combined approach*, i.e. the representation is split into an expansion in terms of spherical harmonics for the long-wavelength part and an expansion in terms of wavelets for the medium- and the high-frequency parts.

In order to consider the Laplacian differential equation our approaches are based on the spherical wavelet theory (Freeden 1999). In the next section some fundamental ideas of spherical signal representation are listed. The two methods mentioned before are explained in detail in the third section. We also outline the application of parameter estimation techniques in order to calculate the unknown model coefficients. Finally in the fourth section the MRRs of both an EGM 96 model data set and Champ in situ potential measurements are presented.

2 Foundations

First we introduce a real-valued signal $x(\mathbf{t})$, which is assumed to be harmonic in the exterior of the Earth, i.e. it fulfills the Laplacian differential equation. In

general the geocentric position vector

$$\mathbf{t} = r\left[\cos\beta\cos\lambda, \cos\beta\sin\lambda, \sin\beta\right]' = r\,\mathbf{r} \tag{1}$$

of any arbitrary observation point $P = P(\mathbf{t})$ may be expressed by means of the spherical coordinates $\lambda =$ longitude, $\beta =$ latitude and $r =$ radial distance from the geocenter; $\mathbf{r}$ denotes the unit vector. The solution of *Dirichlet's problem* for the outer space Ω_R^{ext} of the sphere Ω_R with radius R can be expressed by the series expansion

$$x(\mathbf{t}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} X_{n,m}^R\, Y_{n,m}^R(\mathbf{t}) \tag{2}$$

in terms of the *solid spherical harmonics*

$$Y_{n,m}^R(\mathbf{t}) = \\ = \frac{1}{\sqrt{4\pi R}} \left(\frac{R}{r}\right)^{n+1} \cdot \left\{ \begin{array}{l} P_{n,m}(\sin\beta)\,\cos m\lambda \\ \quad \text{for } m = 0,\ldots,n \\ P_{n,|m|}(\sin\beta)\,\sin|m|\lambda \\ \quad \text{for } m = -n,\ldots,-1 \end{array} \right\}; \tag{3}$$

see e.g. Heiskanen and Moritz (1967, p.34). $P_{n,m}(\cdot)$ are the associated Legendre functions of degree n and order m. Note, that $\overline{\Omega_R^{ext}} = \Omega_R^{ext} \cup \Omega_R$ means the union of the outer space Ω_R^{ext} and the sphere Ω_R. The *Stokes coefficients* $X_{n,m}^R$ are computed by the spherical *Fourier transform*

$$X_{n,m}^R = \int_{\Omega_R} x(\mathbf{t}_q)\, Y_{n,m}^R(\mathbf{t}_q)\, d\omega_R(\mathbf{t}_q)\,, \tag{4}$$

where $d\omega_R(\mathbf{t}) = R^2 \cos\beta d\beta d\lambda$ denotes the surface element on Ω_R; $\mathbf{t}_q \in \Omega_R$. Now we assume, that the signal $x(\mathbf{t})$ is *band-limited*, i.e.

$$X_{n,m}^R = 0 \qquad \text{for} \qquad n > n_3\,. \tag{5}$$

With $(n_3+1)^2 =: \bar{n}$ we introduce the $\bar{n} \times 1$ vectors

$$\boldsymbol{\beta} = \left[X_{0,0}^R, X_{1,-1}^R, \ldots, X_{n_3,n_3}^R \right]', \tag{6}$$

$$\mathbf{z_t} = \left[Y_{0,0}^R(\mathbf{t}), Y_{1,-1}^R(\mathbf{t}), \ldots, Y_{n_3,n_3}^R(\mathbf{t}) \right]' \tag{7}$$

and obtain from Eq. (2) the representation

$$x(\mathbf{t}) = \mathbf{z}_\mathbf{t}'\,\boldsymbol{\beta}\,. \tag{8}$$

The *spherical convolution* of the band-limited function $x(\mathbf{t})$ with a kernel function

$$k^R(\mathbf{t},\mathbf{t}_k) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi R^2} \left(\frac{R^2}{r r_k}\right)^{n+1} K(n)\, P_n(\mathbf{r}'\,\mathbf{r}_k) \tag{9}$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} K(n)\, Y_{n,m}^R(\mathbf{t})\, Y_{n,m}^R(\mathbf{t}_k) \tag{10}$$

is defined as

$$(k^R \star x)(\mathbf{t}_k) = \int_{\Omega_R} k^R(\mathbf{t},\mathbf{t}_k)\, x(\mathbf{t})\, d\omega_R(\mathbf{t}) \tag{11}$$

$$= \sum_{n=0}^{n_3} \sum_{m=-n}^{n} K(n)\, X_{n,m}^R\, Y_{n,m}^R(\mathbf{t}_k)\,. \tag{12}$$

The functions $P_n(\cdot) \equiv P_{n,0}(\cdot)$ are the Legendre polynomials of degree n. The Legendre coefficients $K(n)$ with $n \in \mathbb{N}_0$ reflect the frequency behavior of the kernel. If the kernel function is band-limited in the same manner as the function $x(\mathbf{t})$, i.e.

$$K(n) = 0 \qquad \text{for} \qquad n > n_3\,, \tag{13}$$

we rewrite the Eqs. (10) and (12) in matrix notation

$$k^R(\mathbf{t},\mathbf{t}_k) = \mathbf{z}_{\mathbf{t}_k}'\,\mathbf{K}\,\mathbf{z}_\mathbf{t}\,, \tag{14}$$

$$(k^R \star x)(\mathbf{t}_k) = \mathbf{z}_{\mathbf{t}_k}'\,\mathbf{K}\,\boldsymbol{\beta}\,. \tag{15}$$

If, in addition to (13), the Legendre coefficients $K(n)$ fulfill the condition

$$K(n) > 0 \qquad \text{for} \qquad 0 \le n \le n_3\,, \tag{16}$$

the $\bar{n} \times \bar{n}$ diagonal matrix $\mathbf{K} = \mathrm{diag}(K(0), K(1), K(1), \ldots, K(n_3))$ is positive definite. Next, we introduce the $N \times 1$ vector

$$\mathbf{k_t} = \left[k^R(\mathbf{t},\mathbf{t}_1), k^R(\mathbf{t},\mathbf{t}_2), \ldots, k^R(\mathbf{t},\mathbf{t}_N) \right]' \tag{17}$$

and obtain with Eq. (14) the linear equation system

$$\mathbf{k_t} = \mathbf{Z}\,\mathbf{K}\,\mathbf{z}_\mathbf{t}\,, \tag{18}$$

wherein

$$\mathbf{Z} = \left[\mathbf{z}_{\mathbf{t}_1}, \mathbf{z}_{\mathbf{t}_2}, \ldots, \mathbf{z}_{\mathbf{t}_N} \right]' \tag{19}$$

is an $N \times \bar{n}$ matrix. The $\bar{n}$ components of the vector $\mathbf{z_t}$, Eq. (7), establish a complete orthonormal basis of an $\bar{n}$-dimensional space, denoted by $H_{0,\ldots,n_3}(\Omega_R)$. If for $N \ge \bar{n}$ the matrix $\mathbf{Z}$ is of full column rank, i.e. $\mathrm{rank}\,\mathbf{Z} = \mathrm{rank}\,(\mathbf{ZK}) = \bar{n}$, the altogether N components $k^R(\mathbf{t},\mathbf{t}_k)$ with $k = 1,\ldots,N$ of the vector $\mathbf{k_t}$, Eq. (17), establish another basis of the space $H_{0,\ldots,n_3}(\Omega_R)$. In this case the system

$$S_N(\Omega_R) = \{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_N\} \tag{20}$$

of points $P(\mathbf{t}_k)$ on the sphere Ω_R is called *admissible*. If even the equality $N = \bar{n}$ holds, the matrix $\mathbf{Z}$ is regular and $S_N(\Omega_R)$ is called a *fundamental* point system. In the following we always assume that the point system $S_N(\Omega_R)$ is at least admissible.

We assume now that the Legendre coefficients $K(n) =: K_1(n)$ of a kernel function $k_1^R(\mathbf{t}, \mathbf{t}_k)$ fulfill the conditions (13) and (16). Since both functions $x(\mathbf{t})$ and $k_1^R(\mathbf{t}, \mathbf{t}_k)$ are members of the space $H_{0,\ldots,n_3}(\Omega_R)$, the convolution $(k_1^R \star x)(\mathbf{t})$ is a member of the same space and can be modeled as

$$(k_1^R \star x)(\mathbf{t}) = \mathbf{k}'_{1;\mathbf{t}}\, \mathbf{d}\,, \tag{21}$$

wherein $\mathbf{d} = (d_k)$ is an $N \times 1$ vector of initially unknown coefficients d_k with $k = 1, \ldots, N$. Furthermore, substituting Eq. (18) with $\mathbf{K} =: \mathbf{K}_1$ for $\mathbf{k}_{1;\mathbf{t}}$ yields

$$(k_1^R \star x)(\mathbf{t}) = \mathbf{z}'_{\mathbf{t}}\, \mathbf{K}_1\, \mathbf{Z}'\, \mathbf{d}\,. \tag{22}$$

Comparing the right hand sides of the Eqs. (15) and (22) gives the expression

$$\boldsymbol{\beta} = \mathbf{Z}'\, \mathbf{d}\,. \tag{23}$$

Hence, according to (6), the Stokes coefficients $X_{n,m}$ with $n \leq n_3$ and $-n \leq m \leq n$ are calculable by means of the coefficient vector $\mathbf{d}$. Next, we introduce a second kernel function $k_2^R(\mathbf{t}, \mathbf{t}_k)$. Its Legendre coefficients $K_2(n)$ fulfill certainly the condition (13), but they are not restricted to the condition (16). To be more specific, we allow with

$$K_2(n) \geq 0 \qquad \text{for} \qquad 0 \leq n \leq n_3\,, \tag{24}$$

that the $\overline{n} \times \overline{n}$ matrix $\mathbf{K}_2$ is possibly just positive semidefinite. It follows from Eq. (15) under the consideration of the Eqs. (18) and (23)

$$(k_2^R \star x)(\mathbf{t}) = \mathbf{z}'_{\mathbf{t}}\, \mathbf{K}_2\, \boldsymbol{\beta} = \mathbf{z}'_{\mathbf{t}}\, \mathbf{K}_2\, \mathbf{Z}'\, \mathbf{d} = \mathbf{k}'_{2;\mathbf{t}}\, \mathbf{d}\,. \tag{25}$$

Thus, if the coefficient vector $\mathbf{d}$ is known once, it can be used to calculate both, the Stokes coefficients according to Eq. (23) and any convolution of the function $x(\mathbf{t})$ with kernel functions fulfilling the conditions (13) and (24). As will be explained in the next section, in spherical wavelet theory we identify the kernel functions $k_1^R(\mathbf{t}, \mathbf{t}_k)$ and $k_2^R(\mathbf{t}, \mathbf{t}_k)$ with the scaling function and the wavelet function, respectively.

If the function $x(\mathbf{t})$ is given in the points $P(\mathbf{t}_k)$ of the admissible system $S_N(\Omega_R)$, the components d_k of the coefficient vector $\mathbf{d} = (d_k)$ are computable from

$$d_k = w_k\, x(\mathbf{t}_k)\,, \tag{26}$$

wherein $w_k = w(\mathbf{t}_k)$ are the *integration weights*. A well-known example of an admissible points systems is the *standard longitude-latitude grid*, i.e. the points $P(\mathbf{t}_k) \equiv P(\lambda_i, \beta_j)$ are defined by $\lambda_i = i\frac{2\pi}{2L}$ for $i = 0, \ldots, 2L-1$ and $\beta_j = -\frac{\pi}{2} + j\frac{\pi}{2L}$ for $j = 0, \ldots, 2L$ with $N = (2L+1) \cdot 2L$. The corresponding integration weights $w_k =: w_j$ are merely latitude dependent (Driscoll and Healy 1994). For more details concerning appropriate point grids on the sphere, e.g. the *Reuter grid*, we refer to Freeden et al. (1998, p.171).

Eq. (26) requires that the function $x(\mathbf{t})$ is given in the points $P(\mathbf{t}_k)$ of the admissible system. In the next section we treat a more general approach, namely the computation of the coefficient vector $\mathbf{d}$ with the help of parameter estimation methods.

3 Multi-Resolution Representation

The basic idea of the MRR is to split a given input signal into a smoothed version and a certain number of band-pass signals by *successive low-pass filtering*. In the context of wavelet theory, this procedure consists of the *decomposition* of the signal into wavelet coefficients and the *(re)construction* of the (modified) signal by means of *detail signals*. The latter are the spectral components of the MRR because they are related to certain frequency bands.

3.1 Wavelet-Only Approach

As mentioned before we want to identify the kernel function $k^R(\mathbf{t}, \mathbf{t}_k)$, which fulfills the conditions (13) and (16), with the spherical scaling function. To be more specific we introduce the convolution $(\theta_{I+1}^R \star x)(\mathbf{t})$, wherein $x(\mathbf{t})$ is a band-limited function according to Eq. (5) and

$$\theta_{I+1}^R(\mathbf{t}, \mathbf{t}_k) = \sum_{n=0}^{2^{I+1}-1} \frac{2n+1}{4\pi R^2} \left(\frac{R^2}{r r_k}\right)^{n+1} \Theta_{I+1}(n)\, P_n(\mathbf{r}'\mathbf{r}_k) \tag{27}$$

the band-limited scaling function of highest resolution level $I+1$ with $I \in \mathbb{N}_0$. The Legendre coefficients $\Theta_{I+1}(n)$ fulfill the conditions (13) and (16), i.e.

$$\begin{aligned} &\Theta_{I+1}(n) > 0 \qquad \text{for} \quad n = 0, \ldots, n_3 = 2^{I+1} - 1\,, \\ &\Theta_{I+1}(n) = 0 \qquad \text{for} \quad n > n_3\,; \end{aligned} \tag{28}$$

see e.g. Freeden (1999, p.165). As an example we want to mention the *Shannon scaling function* defined by the Legendre coefficients

$$\Theta_i(n) = \left\{ \begin{array}{ll} 1 & \text{for} \quad n = 0, \ldots, 2^i - 1 \\ 0 & \text{for} \quad n \geq 2^i \end{array} \right\} \tag{29}$$

of resolution level $i \in \mathbb{N}_0$. In order to perform an MRR we introduce the *spherical wavelet function*

$$\psi_i^R(\mathbf{t},\mathbf{t}_k) = \sum_{n=0}^{2^{i+1}-1} \frac{2n+1}{4\pi R^2} \left(\frac{R^2}{r r_k}\right)^{n+1} \Psi_i(n) P_n(\mathbf{r}'\mathbf{r}_k) \tag{30}$$

of resolution level $i \in \mathbb{N}_0$. Its Legendre coefficients $\Psi_i(n)$ are computed via the *two-scale relation*

$$\widetilde{\Psi}_i(n)\Psi_i(n) = \Theta_{i+1}(n) - \Theta_i(n)\,, \tag{31}$$

wherein $\widetilde{\Psi}_i(n)$ are the Legendre coefficients of the so-called *dual spherical wavelet function* $\widetilde{\psi}_i^R(\mathbf{t},\mathbf{t}_k)$; see e.g. Freeden (1999, p.177). Note, that whereas the (primal) spherical wavelet function $\psi_i^R(\mathbf{t},\mathbf{t}_k)$ decomposes a function into its wavelet coefficients, the dual spherical wavelet function $\widetilde{\psi}_i^R(\mathbf{t},\mathbf{t}_k)$ performs the (re)construction (Schmidt 2001, p.273).

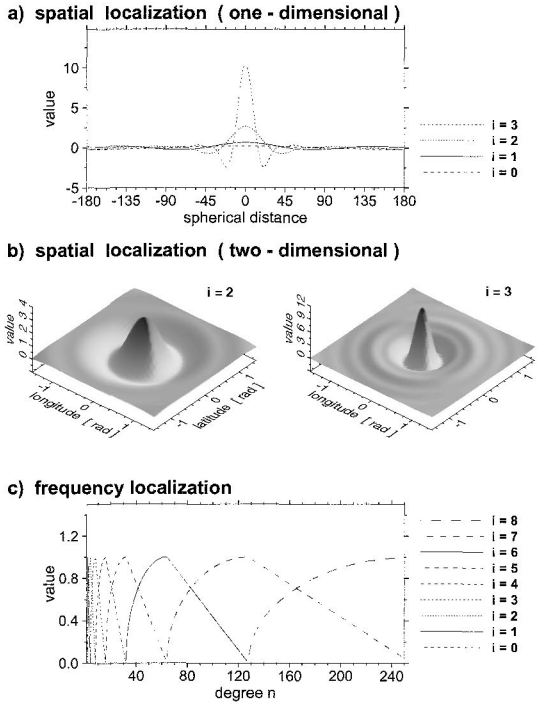


Fig. 1: Smoothed Shannon wavelet with shape parameter $\sigma = 0.5$; see Freeden (1999, p.165), Schmidt et al. (2002).

As an example for a spherical wavelet function with $\widetilde{\psi}_i^R(\mathbf{t},\mathbf{t}_k) = \psi_i^R(\mathbf{t},\mathbf{t}_k)$ the *smoothed Shannon wavelet* is illustrated in Fig. 1. An important advantage of this function is the fact, that it is strictly band-limited, i.e. only a few Legendre coefficients are not equal to zero (Fig. 1c). Hence, according to the condition (24) the spherical wavelet function can be identified with the kernel function $k_2^R(\mathbf{t},\mathbf{t}_k)$.

The Shannon scaling function, Eq. (27) with Eq. (29) for $i = I+1$, is the *reproducing kernel* of the the space $H_{0,\ldots,n_3}(\Omega_R)$ with $n_3 = 2^{I+1}-1$. Hence, it follows with Eq. (21)

$$x(\mathbf{t}) = (\theta_{I+1}^R \star x)(\mathbf{t}) = \boldsymbol{\theta}'_{I+1;\mathbf{t}}\,\mathbf{d} \tag{32}$$

wherein the $N \times 1$ vector $\boldsymbol{\theta}_{I+1;\mathbf{t}}$ is defined analogously to (17). The highest resolution level $I+1$ depends on the maximum degree $n = n_2$ with $n_2 \leq n_3$ of the function $x(\mathbf{t})$ in conformance with Eq. (5); we use the formula $I \geq \log_2(n_2+1)$. Furthermore, it follows from the remarks in the context of Eq. (19), that the number $N =: N_I$ of points $P(\mathbf{t}_{I,k})$ with $k = 1,\ldots,N_I$ of the level-I admissible system $S_{N_I}(\Omega_R)$ is restricted to $N_I \geq 2^{2I+2} =: \overline{n}_I$. In order to estimate the N_I unknown *scaling coefficients* $d_{I,k}$ with $k = 1,\ldots,N_I$ of the $N_I \times 1$ vector $\mathbf{d} =: \mathbf{d}_I$ from Eq. (32), we need the values $x(\mathbf{t}_p) =: x_p$ of the function $x(\mathbf{t})$ in altogether N discrete observation points $P(\mathbf{t}_p)$ with $p = 1,\ldots,N$ and $N \geq N_I$. Note, that henceforth N means the number of observations. However, *geodetic measurements* $y(\mathbf{t}_p) =: y_p$ are always erroneous, i.e. $x_p = y_p + e_p$, wherein $e_p := e(\mathbf{t}_p)$ denotes the measurement error. Under these assumptions Eq. (32) can be rewritten as the *observation equation*

$$y_p + e_p = \boldsymbol{\theta}'_{I+1;p}\,\mathbf{d}_I \tag{33}$$

for a single observation y_p. Note, that probably additional operators, like the Stokes operator, have to be considered in the vector $\boldsymbol{\theta}_{I+1;\mathbf{t}_p} =: \boldsymbol{\theta}_{I+1;p}$, if the observations are functionals of $x(\mathbf{t})$. Introducing the $N \times 1$ vectors $\mathbf{y} = (y_p)$ and $\mathbf{e} = (e_p)$ of the observations and the measurement errors, respectively, the $N \times N_I$ coefficient matrix

$$\mathbf{X} = \left[\,\boldsymbol{\theta}_{I+1;1},\, \boldsymbol{\theta}_{I+1;2},\, \ldots,\, \boldsymbol{\theta}_{I+1;N}\,\right]' \tag{34}$$

and the $N \times N$ covariance matrix $D(\mathbf{y})$ of the observations, the linear model

$$\mathbf{y} + \mathbf{e} = \mathbf{X}\,\mathbf{d}_I \quad \text{with} \quad D(\mathbf{y}) = \sigma^2\,\boldsymbol{\Sigma}_y \tag{35}$$

is established. Herein σ^2 and $\boldsymbol{\Sigma}_y$ are denoted as the variance of unit weight and the matrix of cofactors, respectively. Analog to the matrix $\mathbf{Z}$, Eq. (19), the matrix $\mathbf{X}$ is of rank$\mathbf{X} = \overline{n}_I$, i.e. a rank deficiency of $N_I - \overline{n}_I$ exists. To overcome this rank deficiency problem, the normal equation system, resulting from the linear model (35), might be solved e.g. by means of *Bayesian Inference*, i.e. prior information about the coefficients is introduced; see e.g. Koch (1990). We assume in the following, that the general solution

$$\widehat{\mathbf{d}}_I = \mathbf{B}_I\,\mathbf{y} \tag{36}$$

for the estimator $\widehat{\mathbf{d}}_I$ of the coefficient vector $\mathbf{d}_I$ exists; see e.g. Kusche (2003, p.40); a detailed treatment of regularization problems goes beyond the scope of this paper. Note, that the "hat" in Eq. (36) and in the following always marks an estimation of the corresponding parameter. The estimation of the coefficient vector $\mathbf{d}_I$ means the *initialization step* of the following pyramid algorithm. To be more specific the $(I-i)^{th}$ *pyramid step* consists of the two linear equations

$$\widehat{\mathbf{d}}_i = \mathbf{H}_i\,\widehat{\mathbf{d}}_{i+1} = \mathbf{H}_i\mathbf{H}_{i+1}\ldots\mathbf{H}_{I-1}\,\widehat{\mathbf{d}}_I\,, \tag{37}$$

$$\widehat{\mathbf{c}}_i = \mathbf{\Psi}_i\,\widehat{\mathbf{d}}_i\,, \tag{38}$$

where $\mathbf{H}_i, \mathbf{H}_{i+1}, \ldots, \mathbf{H}_{I-1}$ are *low-pass filter matrices*. In particular $\mathbf{H}_i$ is an $N_i \times N_{i+1}$ matrix, which transforms the $N_{i+1} \times 1$ scaling coefficient vector $\mathbf{d}_{i+1}$ of level $i+1$ into the $N_i \times 1$ scaling coefficient vector $\mathbf{d}_i$ of level i. In Eq. (38) $\mathbf{\Psi}_i$ denotes the $N_i \times N_i$ wavelet matrix of level i, which acts as a *band-pass filter*; $\widehat{\mathbf{c}}_i$ stands for the estimator of the $N_i \times 1$ vector of *level-i wavelet coefficients* $c_{i,k}$ with $k = 1,\ldots,N_i$.

In the *reconstruction process* we start from the wavelet coefficient vector $\mathbf{c}_i$ of level $i \in \{0,\ldots,I\}$ and compute the estimator

$$\widehat{\mathbf{g}}_i = \widetilde{\mathbf{\Psi}}_i\,\widehat{\mathbf{c}}_i \tag{39}$$

of the $M \times 1$ detail signal vector $\mathbf{g}_i$. Herein the $M \times N_i$ dual wavelet matrix $\widetilde{\mathbf{\Psi}}_i$ works as a *band-pass filter*. The elements $g_i(\mathbf{t}_q)$ of the vector $\mathbf{g}_i$ are related to the points $P(\mathbf{t}_q)$ with $\mathbf{t}_q \in \overline{\Omega_R^{ext}}$ and $q = 1,\ldots,M$, which do not necessarily coincide with the points $P(\mathbf{t}_{I,k})$ of the admissible system or the observation sites $P(\mathbf{t}_p)$. Finally the estimated MRR reads

$$\widehat{\mathbf{x}} = \widehat{\mathbf{x}}_0 + \sum_{i=0}^{I} \widehat{\mathbf{g}}_i\,. \tag{40}$$

Herein $\widehat{\mathbf{x}}$ means the $M \times 1$ vector of *filtered* or *predicted* signal values $\widehat{x}(\mathbf{t}_q)$. The components of the vector $\widehat{\mathbf{x}}_0 = (\widehat{x}_0(\mathbf{t}_q))$ are calculated from $\widehat{x}_0(\mathbf{t}_q) = \boldsymbol{\theta}'_{0;q}\,\widehat{\mathbf{d}}_0$. As an example for the signal $x(\mathbf{t})$ we want to mention the *disturbing potential*.

3.2 Combined Approach

In the following we assume that the unknown function $x(\mathbf{t})$, introduced in the previous subsection, can be expressed by means of a series expansion in terms of spherical harmonics (according to Eq. (2), but only until degree $n = n_1$) for the long-wavelength part and a spherical wavelet expansion for the remaining medium- and short-wavelength parts. Hence, the observation equation (33) can be replaced by

$$y_p + e_p = \mathbf{z}'_p\,\boldsymbol{\beta} + \boldsymbol{\theta}'_{I+1,n_1+1;p}\,\mathbf{d}_I\,, \tag{41}$$

wherein with $(n_1+1)^2 =: \bar{n}_1$ the $\bar{n}_1 \times 1$ vectors $\mathbf{z}_{\mathbf{t}_p} =: \mathbf{z}_p$ and $\boldsymbol{\beta}$ are defined by Eqs. (6) and (7) substituting n_1 for n_3. The $N_I \times 1$ vector $\boldsymbol{\theta}_{I+1,n_1+1;p}$ is computed analog to Eq. (32) with the spherical scaling function

$$\theta_{I+1,n_1+1}(\mathbf{t},\mathbf{t}_k) = \sum_{n=n_1+1}^{2^{I+1}-1} \frac{2n+1}{4\pi R^2}\left(\frac{R^2}{r r_k}\right)^{n+1} P_n(\mathbf{r}'\mathbf{r}_k) \tag{42}$$

according to the Eqs. (27) and (29). But in opposite to Eq. (27) the lower summation limit is now set to $n = n_1+1$. With the $N \times u_1$ coefficient matrix $\mathbf{X}_1 \equiv \mathbf{Z}$ according to (19) and the $N \times u_2$ coefficient matrix $\mathbf{X}_2$, analog to (34), we obtain the linear model

$$\mathbf{y} + \mathbf{e} = \mathbf{X}_1\,\boldsymbol{\beta}_1 + \mathbf{X}_2\,\boldsymbol{\beta}_2 \quad \text{with} \quad D(\mathbf{y}) = \sigma^2\boldsymbol{\Sigma}_y \tag{43}$$

of the combined approach, wherein $\boldsymbol{\beta}_1 = (X^R_{n,m})$ is the $u_1 \times 1$ vector of the $u_1 \equiv \bar{n}_1$ (unknown) Stokes coefficients $X^R_{n,m}$ and $\boldsymbol{\beta}_2 \equiv \mathbf{d}_I = (d_{I,k})$ the $u_2 \times 1$ vector of the $u_2 \equiv N_I$ unknown scaling coefficients $d_{I,k}$. Even if the matrix $\mathbf{X}_1$ is of full column rank, the rank of the $N \times u$ matrix $[\mathbf{X}_1, \mathbf{X}_2]$ with $u := u_1 + u_2$ is equal to the rank of the coefficient matrix $\mathbf{X}$ of the wavelet-only model (35). If we assume, that due to the definition (42) of the scaling function, the spherical wavelet model part is independent of the spherical harmonic model part, the normal equation system degenerates into two parts and allows a separate determination of the estimators $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ of the unknown parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. Note, that due to Eq. (23), it is also possible to compute both estimators together from the estimator $\widehat{\mathbf{d}}_I$, Eq. (36), of the wavelet-only model (35). The scaling function $\theta_{I+1,n_1+1}(\mathbf{t},\mathbf{t}_k)$, Eq. (42), fulfills the condition (24) just as the chosen wavelet functions of levels $i = i',\ldots,I$. According to the observation equation (41) the MRR (40) has to be replaced by

$$\widehat{\mathbf{x}} = \widehat{\mathbf{x}}_{0,\ldots,n_1} + \sum_{i=i'}^{I} \widehat{\mathbf{g}}_i\,, \tag{44}$$

wherein $\widehat{\mathbf{x}}_{0,\ldots,n_1}$ means the $M \times 1$ vector of the estimated spherical harmonic signal values computed in the points $P(\mathbf{t}_q)$.

4 Numerical Examples

In the first example the combined approach is applied to *gravity anomalies of EGM 96*. In order to demonstrate the method and to keep the numerical efforts

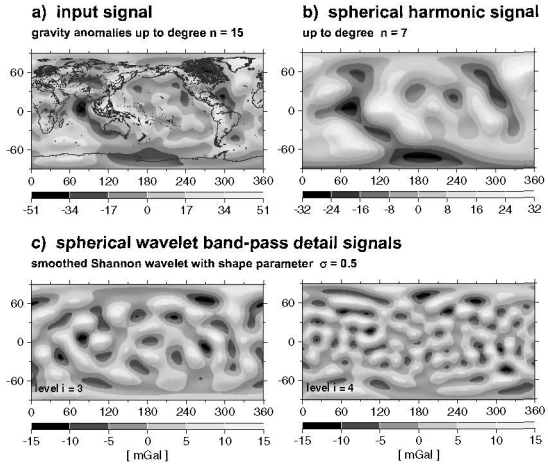


Fig. 2: Gravity anomalies of EGM 96 up to degree $n_3 = 15$ and results of the combined estimation.

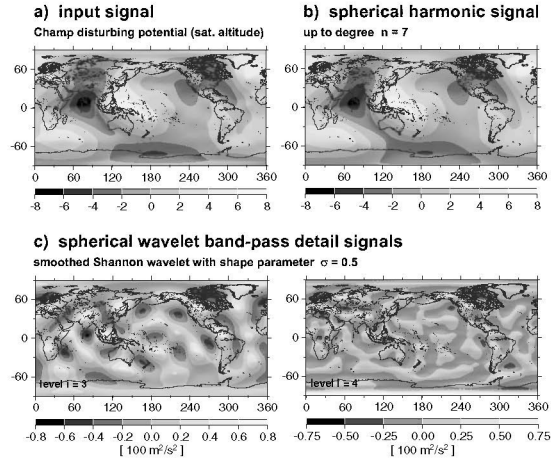


Fig. 3: Champ disturbing potential decomposition at satellite altitude (450 km).

low we choose a maximum degree $n_2 = 15$ for the input signal (Fig. 2a). With $n_1 = 7$ the series expansion in terms of spherical harmonics contains $u_1 = 64$ unknown Stokes coefficients. As spherical wavelet function we choose the smoothed Shannon wavelet shown in Fig. 1. Since, according to Fig. 1c the detail signals of the levels $i = 3 = i'$ and $i = 4 = I$ have to be considered, the MRR (44) reduces to

$$\widehat{\mathbf{x}} = \widehat{\mathbf{x}}_{0,\ldots,7} + \widehat{\mathbf{g}}_3 + \widehat{\mathbf{g}}_4 \,. \tag{45}$$

As admissible point system $S_{N_4}(\Omega_R)$ we use the Reuter grid, mentioned before. After computing the estimators $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2 \equiv \widehat{\mathbf{d}}_4$ from the linear model (43) in the initialization step, the following pyramid steps are performed according to the Eqs. (37) to (39). The Figs. 2b, c display the values of the three estimated vectors $\widehat{\mathbf{x}}_{0,\ldots,7}$, $\widehat{\mathbf{g}}_3$ and $\widehat{\mathbf{g}}_4$. According to Eq. (45) the sum of these vectors means an estimation of the MRR of the input gravity anomalies visualized in Fig. 2a. The deviations (not shown here) between the input signal and the MRR are in the range of the computational accuracy. Due to the ability of wavelet functions to localize, the block matrix $\mathbf{X}_2' \boldsymbol{\Sigma}_y^{-1} \mathbf{X}_2$ is, depending on the chosen wavelet function, possibly close to be of sparse form. The localizing property gives also reason to apply this procedure to regional data sets. Numerical investigations concerning this topic will be presented in an forthcoming paper of the authors.

In the second example the same procedure is applied to *Champ in situ disturbing potential values* in satellite altitude (Fig. 3a), computed from Champ data by means of the energy conservation principle. Again the sum of the estimated spherical harmonic signal $\widehat{\mathbf{x}}_{0,\ldots,7}$ (Fig. 3b) and the estimated detail signals $\widehat{\mathbf{g}}_3$ and $\widehat{\mathbf{g}}_4$ (Fig. 3c) yields the MRR of the input signal (Fig. 3a). But this time the deviations (not shown here) between the input signal and the MRR consist of both stochastic measurement errors and the detail signals of resolution levels higher than $I = 4$. The rms value of the deviations amounts less than 0.1 m^2/s^2.

References

Beylkin, G. and R. Cramer (2002): Towards Multiresolution Estimation and Effi cient Representation of Gravitational Fields. Celestial Mechanics and Dynamical Astronomy, 84, 87-104

Driscoll, J.R. and R.M. Healy (1994): Computing Fourier Transforms and Convolutions on the 2-Sphere. Adv. Appl. Math., 15, 202-250

Freeden, W. (1999): Multiscale Modelling of Spaceborne Geodata. Teubner, Stuttgart

Freeden, W., Gervens, T. and M. Schreiner (1998): Constructive Approximation on the Sphere (With Applications to Geomathematics). Clarendon Press, Oxford

Heiskanen, W. and H. Moritz (1967): Physical Geodesy. Freeman, San Francisco

Koch, K.R. (1990): Bayesian Inference with Geodetic Applications. Springer, Berlin

Kusche, J. (2002): Inverse Probleme bei der Gravitationsfeldbestimmung mittels SST- und SGG-Satellitenmissionen. Postdoctoral thesis, German Geodetic Commission, München

Schmidt, M. (2001): Grundprinzipien der Wavelet-Analyse und Anwendungen in der Geodäsie. Postdoctoral thesis, Shaker, Aachen

Schmidt, M., Fabert, O. and C.K. Shum (2002): Multi-Resolution Representation of the Gravity Field Using Spherical Wavelets. Weikko A. Heiskanen Symposium, The Ohio State University, Columbus

Stokes Integration versus Wavelet Techniques for Regional Geoid Modelling

Markus Roland, Heiner Denker
Institut für Erdmessung, Universität Hannover
Schneiderberg 50, D-30167 Hannover, Germany
Email: roland@ife.uni-hannover.de, Fax: +49 511 762 4006

Abstract. For the computation of high resolution regional geoid models, gravity and terrain data in connection with a global geopotential model play a very important role. The data sets are usually combined in a remove-restore procedure. In many cases, the transformation from gravity anomalies to geoid undulations is done using Stokes's integration kernel or a modified integration kernel, e.g., based on the spectral combination technique. Least squares collocation may be used for this task as well, but for continental-scale computations the integration techniques are often preferred due to their high computational efficiency.

Besides the classical integration techniques, the wavelet technique is investigated in this contribution. The wavelet technique also uses residual gravity field quantities in a remove-restore procedure. However, the computations are carried out in two steps. The first step consists of a convolution of the residual gravity data with several wavelet functions, being contracted or dilated variants of one prototype ("mother") wavelet function. This leads to a decomposition of the whole spectrum of the original data into a set of filtered detail signals with unique spatial resolution. This type of space and frequency analysis is called multi-scale analysis (MSA). The second step then convolves the residual gravity details with an integration kernel (e.g., Stokes) and leads to corresponding geoid undulations. The second step, applied to every decomposed detail (scale) of the original data, corresponds to the classical integration techniques.

In this contribution, both the classical integration and spherical wavelet techniques are applied using Europe as a test area. The differences in methodology and numerical performance of both techniques are investigated. Finally, the results are evaluated by independent GPS and levelling control points.

Keywords. Stokes, multi-scale analysis, spherical wavelets

1 Introduction

The determination of high resolution regional geoid models is accomplished by the combination of various gravity field observations, all of them having unique properties in terms of data coverage and spectral behavior. Usually a global gravity field model, terrestrial gravity data and terrain data are combined in a remove-restore procedure, where the long wavelength content of the global model and the short wavelength content of a topographic model are subtracted from the surface gravity data. The resulting residual gravity anomalies contain mainly the medium wavelengths of the gravity spectrum and have small values.

The conversion of these residual gravity anomalies to geoid undulations can be realized by different techniques. One method is least squares collocation, but for continental-scale computations a second method, the integration technique, is preferred due to its higher computational efficiency. The integration is done using Stokes's integration kernel or a modified integration kernel, e.g., based on the spectral combination technique. A third technique for geoid computation is the wavelet modelling, see Freeden and Schneider (1998). It is also based on Stokes's formula and therefore partly comparable to the integration technique.

In this paper we will discuss and compare the methodology of both the integration and the spherical wavelet technique. Results are presented for both methods using a $5' \times 5'$ gravity anomaly grid from the European Geoid Project, Denker and Torge (1998). Finally the pros and cons of both methods are outlined and a summary of the comparisons is given.

2 Methodology

2.1 Stokes Integration

The classical way to transform gravity data collected at the surface of the Earth into geoid undulations N was found by Stokes (1849). His formula performs

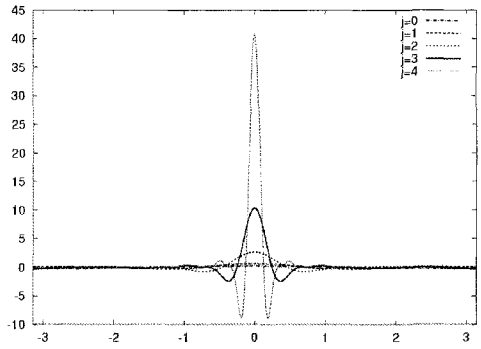


Fig. 1. Smoothed Shannon wavelet function $\Psi_j(\psi)$ for scales $j = 0 \ldots 4$

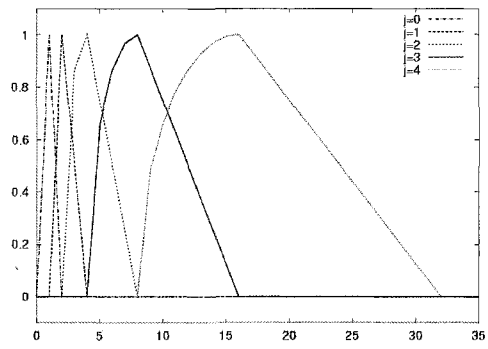


Fig. 2. Legendre coefficients $\Psi_j^\wedge(\ell)$ of smoothed Shannon wavelet for scales $j = 0 \ldots 4$

the integration of gravity anomalies Δg distributed over the whole sphere using the Stokes function $S(\psi)$ as integration kernel, which acts as a weighting function for the anomalies.

$$N = \frac{R}{4\pi\gamma} \int_\Omega S(\psi)\, \Delta g\, d\omega \tag{1}$$

In (1) ψ is the spherical distance between the computation point and the surface element $d\omega$ with the gravity anomaly Δg. Furthermore, γ is the mean normal gravity value over the Earth of radius R. The integration over the whole sphere Ω can be regarded as a spherical convolution of the function Δg with the kernel S, see Freeden et al. (1998). Using this relation equation (1) becomes

$$N = \frac{R}{4\pi\gamma}(S * \Delta g). \tag{2}$$

This convolution can be computed fast and exactly via a 1D Fourier transform, see Haagmans et al. (1993). The Stokes function $S(\psi)$ depends only on the spherical distance ψ and can be computed via a closed formula. The drawback of this function is its infiniteness for $\psi = 0$, so that a special treatment of the neighborhood of the computation point is necessary.

When using residual data, the computation over the whole sphere can be limited to a bounded region under the assumption of zero-values outside this region, e.g. Torge (2001).

2.2 Wavelet Modelling

Wavelets are applied in a wide range of different branches. The common basis for all of them is the wavelet function, which is defined via the contraction and translation of one prototype "mother" wavelet. Various wavelet techniques were developed, e.g., through different realizations of the contraction, or the domain of the function. For instance, there exist discrete and continuous wavelet transforms as well as 1- or 2-dimensional or spherical wavelets, see Grabs (1995); Jawerth and Sweldens (1994); Daubechies (1992); Liu and Sideris (2003); Windheuser (1995). In the spherical case, the translation of the "mother" wavelet is replaced by a rotation.

The modelling of the geoid from gravimetric data by wavelets can be achieved by a multi-scale analysis (MSA), see Freeden et al. (1998). Here the contraction is done in fixed steps resulting in discrete scales j. Furthermore, spherical wavelets, depending only on the spherical distance ψ, are used because the spherical shape of the Earth has to be considered for continental scale computations. The spherical wavelet functions are defined by the Legendre series

$$\Psi_j(\psi) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} \Psi_j^\wedge(\ell) P_\ell(\cos\psi), \tag{3}$$

where the Legendre coefficients $\Psi_j^\wedge(\ell)$ are connected to the generating "mother" wavelet. An example of a wavelet function and its corresponding Legendre coefficients are illustrated in Figs. 1 and 2.

The wavelet modelling by MSA consists of two steps. The first step is the decomposition of the source data, or generally a function F, into its wavelet transforms of scales j

$$\mathrm{WT}\{F\}_j = (\Psi_j * F) = \int_\Omega \Psi_j\, F\, d\omega. \tag{4}$$

This step can be interpreted as a bandpass-filter. The function F is split up into several spectral bands. The bandwidth of each scale j is given by the range of

degrees ℓ where the Legendre coefficients $\hat{\Psi_j}(\ell)$ are unequal 0, see Fig. 2.

The second step of the MSA is the reconstruction of the wavelet transform $\mathrm{WT}\{F\}_j$ to the detail signal F_j of scale j. The summation of all detail signals yields the original function F.

$$\begin{aligned} F &= \sum_{j=0}^{\infty} \left(\mathrm{WT}\{F\}_j * \tilde{\Psi}_j \right) \\ &= \sum_{j=0}^{\infty} \int_{\Omega} \mathrm{WT}\{F\}_j \; \tilde{\Psi}_j \, d\omega = \sum_{j=0}^{\infty} F_j \end{aligned} \tag{5}$$

In the reconstruction, the dual wavelet function $\tilde{\Psi}$ is used, which relates to Ψ via the "refinement equation". In case of the P-scale discrete wavelets used here, Ψ and $\tilde{\Psi}$ are identical, for further details see Freeden et al. (1998).

With the help of the equations (4) and (5) one can decompose a function F into its wavelet transform and exactly reconstruct the same function F from this transform. If we insert the gravity anomalies Δg for function F and put (5) in (1), we get

$$\begin{aligned} N &= \frac{R}{4\pi\gamma} \sum_{j=0}^{\infty} \left(\mathrm{WT}\{\Delta g\}_j * \left(S * \tilde{\Psi} \right)_j \right) \\ &= \frac{R}{4\pi\gamma} \sum_{j=0}^{\infty} \int_{\Omega} \mathrm{WT}\{\Delta g\}_j \left(S * \tilde{\Psi} \right)_j \, d\omega, \end{aligned} \tag{6}$$

a reconstruction/transformation formula for the geoid, see Schmidt (2001) and Schmidt et al. (2002). The details of the geoid undulation are determined by a convolution of the wavelet transform of the gravity anomalies with an integral kernel, which itself results from a convolution of the Stokes's function with the wavelet functions of scale j. The integral kernel is illustrated in Fig. 3. Note that it is finite for small distances ψ, thus no inner-zone correction is necessary.

Both steps of the MSA are realized by a spherical convolution. As shown in Sect. 2.1 and equations (4)–(6), the convolution is equivalent to an integration over the data region and can be computed exactly by a 1D Fourier transform. The necessary kernels Ψ_j and $(S * \tilde{\Psi})_j$ have to be computed via an infinite Legendre series, see (3). For band-limited wavelet functions, where the Legendre coefficients are unequal 0 only for a bounded range of degrees ℓ, the summation can be limited to $2^{j+1} - 1$. But, nevertheless, the exact computation for every ψ is very time-consuming, thus a tabulation of the kernels is useful.

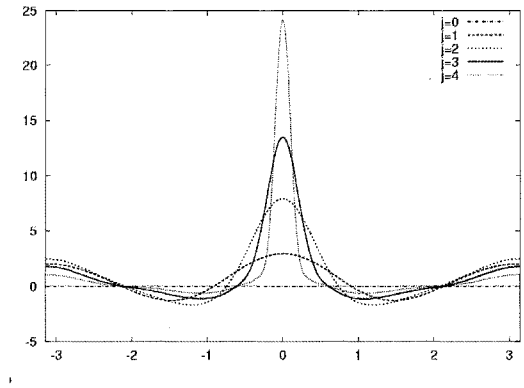


Fig. 3. Integral kernel $(S * \tilde{\Psi})_j(\psi)$ for smoothed Shannon wavelet and scales $j = 0 \ldots 4$

Another problem arises from the infinite summation in the reconstruction formula (6). The total signal is recovered by the summation of *all* detail signals. As the spectral content of the data is limited, the detail signal will vanish for some j, so a truncation is possible. For practical computations the maximum scale is given by the sampling theorem. If the bandwidth of the wavelet function is smaller than twice the resolution of the data, aliasing effects will occur. Consequently, the truncation scale $j_{\max}$ is given by the chosen wavelet function and the data resolution.

2.3 Comparison of Both Methods

Both methods, Stokes integration and wavelet modelling, are based on Stokes's formula. The major difference in methodology consists of the different handling in the frequency domain, leading to a different computational effort.

More precisely, the integration technique handles the whole bandwidth of the residual gravity anomalies at once, so only one convolution is necessary. The wavelet technique first splits the data into several spectral bands by a bandpass-filter and then determines the geoid by handling each spectral band separately. Therefore two convolutions per spectral band are necessary, whereby the second convolution is comparable to the Stokes integration.

3 Results

For the application of both methods, $5' \times 5'$ terrestrial gravity anomalies from the European Geoid Project, see Denker and Torge (1998), were used. The long wavelength part of the gravity field was subtracted using the EGM96 model up to degree and order 360. Furthermore, we subtracted the influence of the topography, determined by a digital terrain model, leading to residual gravity anomalies. The

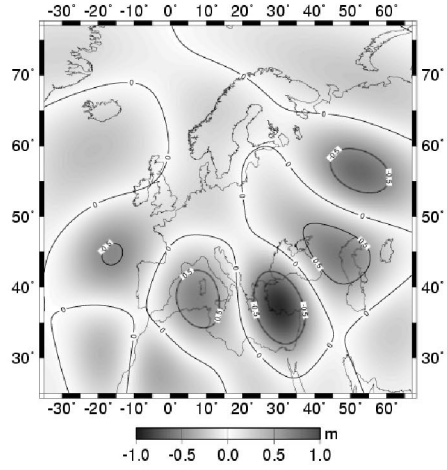


Fig. 4. Detail signal N_4

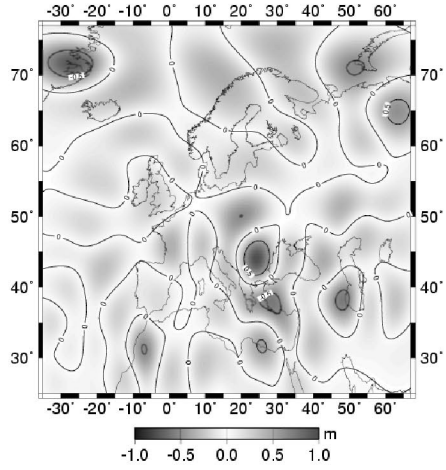


Fig. 5. Detail signal N_5

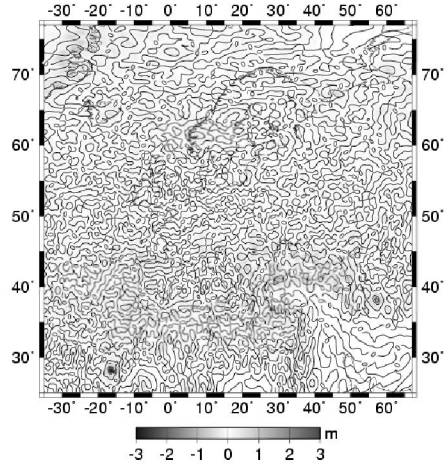


Fig. 6. Detail signal N_8

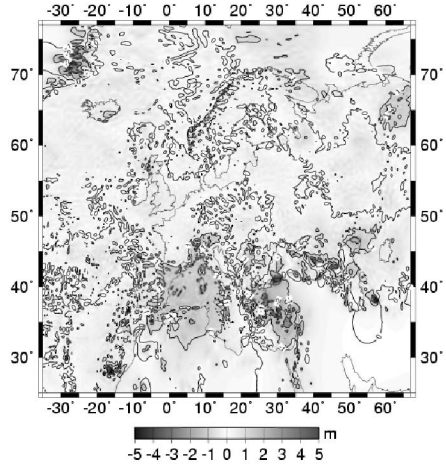


Fig. 7. Sum of detail signals N_j of scales $j = 1 \ldots 11$

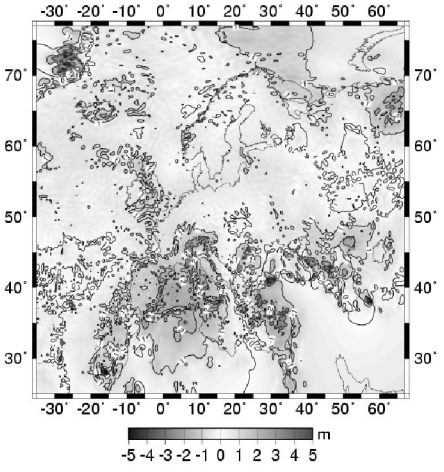


Fig. 8. Geoid undulations determined by Stokes integration

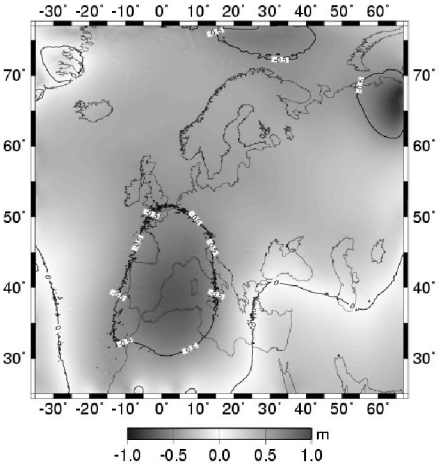


Fig. 9. Difference between Wavelet and Stokes geoid

computation area can be seen in Figs. 4–9.

All results presented here are based on the smoothed Shannon wavelet function with a smoothing factor of $h = 0.5$, illustrated in Fig. 1. The MSA was performed up to a maximum scale of $j = 11$.

In Figs. 4–6 the detail signals, determined by the wavelet technique, are shown for scales 4, 5 and 8 as examples. The figures show a decrease of the wavelengths with increasing scale and a varying amplitude between the scales, which can be attributed to the different energy in the spectral bands.

The sum of all detail signals from scale 1 to 11 is depicted in Fig. 7. This is the final result of the wavelet modelling. The sum contains all degrees up to ℓ=2048. A summation up to scale 12 would contain degrees up to ℓ=4096, but the 5′ data resolution only allows a computation up to ℓ=2160, otherwise aliasing effects will occur. This leads to an approximation error, i.e, a loss of spectral content, because the short wavelengths (ℓ=2049–2160) of the residual gravity anomalies are missing. The differences between the wavelet result in Fig. 7 and the solution by classical Stokes integration in Fig. 8 are illustrated in Fig. 9. The differences show a long wavelength characteristics and have a RMS of 37.1 cm. A possible cause of this effect may be a long wavelength effect in the anomalies reconstructed by the wavelets. This is indicated by the geoid undulations computed from the gravity anomaly approximation error, see Fig. 10, which have a similar long wavelength pattern and a similar RMS of 36.6 cm.

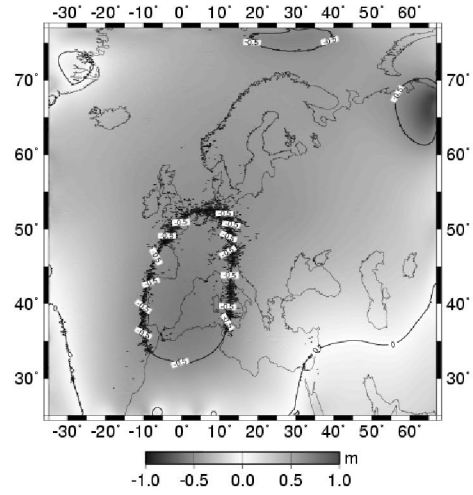


Fig. 10. Geoid undulations from gravity anomaly approximation error

The evaluation of the final results by 166 independent GPS and levelling control points of the UELN (United European Levelling Net), see Ihde et al. (2000), leads to similar results for the Stokes and wavelet method. The bias corrected RMS of the differences between GPS/levelling and the computed geoid undulations are listed in Table 1. The RMS of the wavelet technique is slightly better than that of the Stokes integration.

Table 1. RMS of differences between GPS/levelling and the computed geoid undulations. Units are cm.

Method	RMS
Stokes Integration	49.2
Wavelet Modelling	41.0

4 Summary

In this paper we have shown the similarities but also the differences in methodology of Stokes integration and wavelet modelling in transferring gravity data to geoid undulations.

The main similarity is that both techniques are based on Stokes's formula. From the differences between both methods arise the pros and cons of each technique. The first difference is the computational effort. The wavelet technique takes approx. $2j$ times more time than Stokes. The efficiency may be improved by reducing the integration limits, adapting the localization property of the wavelet function used. The second difference is related to the properties of the kernels. Stokes's kernel is defined via a closed formula, but is infinite at $\psi = 0$, requiring an inner-zone correction. The wavelet kernels are defined by a series expansion. The use of band-limited wavelets and accurate tabulating is necessary to ensure exact and efficient computations. Furthermore, one gains advantage of the finiteness of the wavelet kernels at $\psi = 0$, making further inner zone corrections unnecessary.

A problematic aspect of the wavelet method is its truncation error, arising from the finite summation of the details. This can be reduced by a balanced relation between data resolution and spectral content of the residuals. When raising the data resolution the aliasing effects will be postponed to higher scales, where almost no signal is contained in the residuals due to the terrain reduction. Thus the truncation error will decrease.

The most important advantage of MSA is its great analysis potential, which is, e.g., very useful for the determination of ideal combination solutions from

global gravity field models and terrestrial gravity data.

References

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics.

Denker, H. and Torge, W. (1998). The European Gravimetric Quasigeoid EGG97 – An IAG supported continental enterprise. In Forsberg, R., Feissel, M., and Dietrich, R., editors, *Geodesy on the Move – Gravity, Geoid, Geodynamics, and Antarctica, IAG Symposia*, volume 119, pages 249–254. Springer Verlag.

Freeden, W., Grevens, T., and Schreiner, M. (1998). *Constructive Approximation on the Sphere with Application to Geomathematics*. Clarendon Press, Oxford.

Freeden, W. and Schneider, F. (1998). An integrated wavelet concept of physical geodesy. *Journal of Geodesy*, 72(5):259–281.

Grabs, A. (1995). An Introduction to Wavelets. *IEEE Computational Science and Engineering*, 2(2):50–61. `http://www.amara.com/ftpstuff/IEEEwavelet.pdf`.

Haagmans, R., de Min, E., and von Geldern, M. (1993). Fast evaluation of convolution integrals on the sphere using 1D FFT, and a comparison with existing methods for Stokes' integral. *manuscripta geodaetica*, 18(5):227–241.

Ihde, J., Adam, J., Gurtner, W., Harsson, B. G., Sacher, M., Schlüter, W., and Wöppelmann, G. (2000). The Height Solution of the European Vertical Reference Network (EUVN). Online-Paper `ftp://ftp.leipzig.ifag.de/pub/euvn/EUVN-Height-Solution.pdf`.

Jawerth, B. and Sweldens, W. (1994). An overview of wavelet based multiresolution analyses. *SIAM Rev.*, 36(3):377–412. `http://cm.bell-labs.com/who/wim/papers/overview.ps.gz`.

Liu, Q. and Sideris, M. (2003). Wavelet evaluation of the Stokes and Vening Meinesz integrals. *Journal of Geodesy*, 77(5-6):345–356.

Schmidt, M. (2001). *Grundprinzipien der Wavelet-Analyse und Anwendung in der Geodäsie*. Shaker, Aachen.

Schmidt, M., Martinez, W., and Florez, J. (2002). General Scheme for the Computation of Regional Geoid Undulations Using Spherical Wavelets. In Drewes, H., Dodson, A., Fortes, L., Sánchez, L., and Sandoval, P., editors, *Vertical Reference Systems, IAG Symposia*, volume 124, pages 209–214. Springer Verlag.

Stokes, G. G. (1849). On the variation of gravity on the surface of the earth. *Trans. Cambridge Phil. Soc.*, 8:672–695.

Torge, W. (2001). *Geodesy*. de Gruyter, Berlin, New York, 3rd completly rev. and extended edition.

Windheuser, U. (1995). *Sphärische Wavelets: Theorie und Anwendung in der Physikalischen Geodäsie*. PhD thesis, Fachbereich Mathematik, Universität Kaiserslautern.

A PRELIMINARY GEOID MODEL FOR AFRICA

C L Merry, School of Architecture, Planning & Geomatics, University of Cape Town, South Africa

D Blitzkow, Escola Politécnica, Universidade de São Paulo, São Paulo, Brazil

H Abd-Elmotaal, Department of Civil Engineering, Minia University, Minia, Egypt

H H Fashir, Department of Surveying, University of Khartoum, Khartoum, Sudan

S John, University College, Dar es Salaam, Tanzania

F Podmore, Department of Surveying, University of Zimbabwe, Harare, Zimbabwe

J D Fairhead, School of Earth Science, University of Leeds, Leeds, United Kingdom

Abstract. Precise geoid models are essential for the conversion of GPS-derived heights to heights above sea level. Such a model is under development for the continent of Africa, as part of the African Geoid Project. A uniform 5' grid of gravity anomalies has been derived from terrestrial gravity data and has been combined with a 5' grid for the marine areas derived from satellite altimetry. The combined data set has been used with the EGM96 geopotential model in a remove-restore process to compute the geoid using two-dimensional convolution. The final result is a 5' grid of geoidal heights covering the land mass of Africa. There are significant gaps in the available terrestrial gravity data - these gaps, coupled with the effects of errors in the DEM used for calculating the G_1 term and in interpolating gravity anomalies, mean that the accuracy of this geoid model will be variable and generally less than desirable. Nevertheless, comparison with GPS/levelling data covering a small part of South Africa shows an RMS agreement of better than 10cm. Over a larger region (all of Egypt) the agreement is less satisfactory.

Keywords. Geoid, quasi-geoid, convolution, G_1, Molodensky, Africa

1 Introduction

Precise geoid models are now an essential prerequisite for the successful conversion of GPS-derived ellipsoidal heights to orthometric heights. There are many examples of such geoid models in the developed world (Roman and Smith 2000; Featherstone et al. 2001). Africa has always lagged behind other continents in terms of geodetic research, and the quest for a precise geoid model is no exception. Although much work has been done on developing geoid models for Africa (Obenson 1985; Merry and van Gysen 1987; Gachari and Olliver 1998; Fashir et al. 1998; Benahmed Daho and Kahlouche 2000; Amod and Merry 2002a), these models cannot be said to be either precise or complete. More recently, a collaboration between African geodesists initiated by the Committee for Developing Countries of the International Association of Geodesy has led to a project for the computation of the geoid in Africa – the African Geoid Project (Merry and Blitzkow 2001).

This project will attempt to address some of the challenges facing geoid modelling which are specific to Africa:

- sparcity of data, and difficulty in obtaining data
- quality of data - lack of uniformity in datums
- quality of DEM's and density of DEM's

2 Mathematical Models

The basis for the geoid computation is the so-called "remove-restore" technique (Forsberg and Tscherning 1981), where the contribution of the long wavelength component is removed from the gravity anomalies, the anomalies are processed to obtain geoidal heights, and the long wavelength contribution is restored to the heights. The geoidal height N is derived from the height anomaly ζ using the expression (Rapp 1997):

$$N = \zeta + \frac{\bar{g} - \bar{\gamma}}{\bar{\gamma}} H \qquad (1)$$

where $\bar{g}$ and $\bar{\gamma}$ are mean actual and mean normal gravity respectively, and H is the orthometric height. Rapp (1997) has generated a spherical harmonic representation of this correction term. The height anomaly ζ is given by the summation of four terms:

$$\zeta = \zeta_L + \zeta_S + \zeta_{G1} + \zeta_I \qquad (2)$$

where ζ_L is the long wavelength component, ζ_S is the short wavelength component, ζ_{G1} is the contribution of the Molodensky G_1 term and ζ_I is the inner zone contribution. The long wavelength component, ζ_L, is given by the standard spherical harmonic expansion, as can be found in Heiskanen and Moritz (1967). For this synthesis the EGM96 geopotential model, complete to degree 360 (Lemoine et al. 1998) was used.

The short wavelength component is given by the convolution of the reduced gravity anomaly, Δg_r with Stokes' function, $S(\psi)$:

$$\zeta_S = \frac{R\Delta\phi\Delta\lambda}{4\pi\gamma}\left[S(\psi) * \Delta g_r \cos\phi\right] \qquad (3)$$

Here, R is a mean Earth radius, and $\Delta\phi$ and $\Delta\lambda$ are the grid intervals in latitude and longitude directions for a regular grid of reduced gravity anomalies. These anomalies are created by subtracting the contribution of the EGM96 geopotential model from observed free-air gravity anomalies Δg_f:

$$\Delta g_r = \Delta g_f - \Delta g_L \qquad (4)$$

where Δg_L is the long wavelength contribution generated from a spherical harmonic expansion using the EGM96 coefficients.

The contribution of the G_1 term involves a convolution similar to that used for ζ_S, where the G_1 anomaly is substituted for the reduced anomaly:

$$\zeta_{G1} = \frac{R.\Delta\phi.\Delta\lambda}{4\pi\gamma}\left[S(\psi) * G_1 \cos\phi\right] \qquad (5)$$

The G_1 anomaly is determined from a convolution of free-air anomalies and heights with a kernel consisting of the inverse distance cubed:

$$G_1 = \frac{\Delta\phi\Delta\lambda}{2\pi}\left[(h\Delta g_f) * \frac{1}{\ell^3} - h\left(\Delta g_f * \frac{1}{\ell^3}\right)\right] \qquad (6)$$

The inner zone contribution uses the standard linear approximation for the gravity anomaly field in this inner zone (Heiskanen and Moritz, 1967).

3 Data

3.1 Harmonic Coefficients

The basis for the gravimetric geoid model is a set of spherical harmonic coefficients, derived from satellite and terrestrial data. The most recent and comprehensive data set is the Earth Gravity Model 1996 - EGM96 (Lemoine et al. 1998). This consists of a set of coefficients complete to degree and order 360, with an effective wavelength of 1°.

3.2 Gravity Anomalies

The detailed structure of the quasi-geoid is resolved using a set of free-air gravity anomalies. Data on both land and sea are required. The land data for southern Africa are taken from the UCT gravity database (van Gysen and Merry 1987); land data for the rest of Africa come from the African Gravity Project (Fairhead et al. 1988). Figure 1 shows the land gravity data used in this computation.

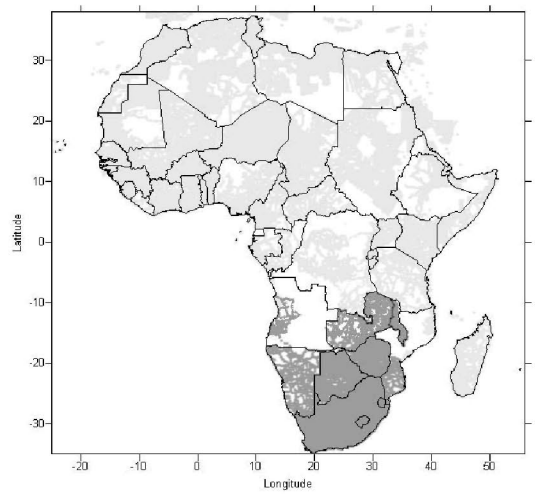


Fig. 1 Point Gravity Values - Africa
(source: - Getech - UCT)

These data are irregularly distributed and were interpolated onto a regular grid before they could be used for two-dimensional convolution. The inter-

polation process involved interpolating simple Bouguer gravity anomalies onto a 5' grid using Kriging and then converting these anomalies to free-air anomalies using a 5' DEM.

Sea gravimetry data are notoriously unreliable and in the case of Africa are very sparse, and free-air gravity anomalies predicted from satellite altimetry were used instead. The 2' x 2' KMS99 data set, produced by the Danish National Survey and Cadastre (Andersen and Knutsen 1998) from ERS-1 and GEOSAT satellite altimetry was used to interpolate a 5'x5' grid for the sea areas.

The land and sea data were combined into one set. Where gaps in the gridded data still existed (mainly in Angola, Congo, Libya and Mozambique), they were filled using EGM96 gravity anomalies.

In order to reduce computational effort and to avoid computing a geoid model over a large part of the south Atlantic ocean, the data sets (and the computations) were split into two parts, with a 4° overlap in latitude. The 5'x5' gravity anomaly grid covered the areas

Southern Africa: $-37° \le \phi \le 4°$; $5° \le \lambda \le 53°$
Northern Africa: $0° \le \phi \le 39°$; $-20° \le \lambda \le 53°$

A free-air anomaly grid is also required for the purpose of computing the G_1 anomaly. In this case as dense a grid as possible is desirable. However, the calculation is numerically intensive, and the original gravity data spacing is no better than 5', so a compromise had to be reached between computational effort, resolution and accuracy. The compromise chosen was to use a 1' grid for the calculation of the G_1 anomaly. Consequently, a 1'x1' free-air anomaly data set was created by interpolation from the 5'x5' data set.

3.3 Digital Elevation Model

Height data form an important part of the quasi-geoid computation. Not only is the DEM used to aid in the interpolation process (converting interpolated Bouguer anomalies into free-air anomalies), but it is also essential for the computation of the Molodensky G_1 term. Africa is poorly served by DEM's. The GLOBE model (Hastings and Dunbar 1998) is the most comprehensive, but is burdened with errors (Berry et al. 1999), which are not insignificant in the context of quasi-geoid modelling (Merry 2003). For lack of alternatives, the GLOBE DEM has been used in this project. The 30"x30" data set has been decimated to 1'x1' data sets (needed for the G_1 term) and 5'x5' data sets (needed for the prediction of free-air anomalies), for each of southern and northern Africa.

4 Data Processing and Results

4.1 Long Wavelength Geopotential Field

These components, Δg_L and ζ_L, were computed from the EGM96 harmonic coefficients on 5'x5' grids for southern and northern Africa. A correction of -0.53 m has been added, to refer the quasi-geoidal heights to the WGS84 ellipsoid.

4.2 Residual Quasi-Geoid

The calculation of the short wavelength component, ζ_S, and the inner zone contribution, ζ_I, were computed at the same time using the two-dimensional convolution operator together with residual gravity anomalies. The spherical Stokes' kernel described by Strang van Hees (1990) was used. In order to compensate for the variable $cos\phi$ term in the kernel, the convolution was repeatedly applied for each parallel of latitude. Although convolution is computationally less efficient than FFT, in this implementation a 4° x 4° spatial window was used (akin to an integration distance of 2° in the numerical integration of Stokes' formula), which improves the efficiency of convolution. The edge effect was eliminated by discarding a border two degrees wide around each region. The computed residual quasi-geoid models encompass the two regions:

Southern Africa: $-35° \le \phi \le 2°$; $7° \le \lambda \le 51°$
Northern Africa: $2° \le \phi \le 37°$; $-18° \le \lambda \le 51°$

The inner zone contribution ζ_I was calculated on a 5' grid and was then added to the short wavelength component. The combined result represents that contribution of the gravity field not covered by the EGM96 geopotential model. The statistics of this contribution, together with those of EGM96 and the G_1 and N-ζ terms, are summarised in Table 1.

Table 1. Contributions to African Geoid

Contribution	Min. (m)	Max. (m)	Mean (m)	RMS (m)
EGM96 ζ	-52.68	+56.67	+10.84	24.50
Residual ζ	-4.88	+7.51	+0.11	0.67
G_1 ζ	-0.19	+0.78	+0.03	0.07
N - ζ	-0.89	+0.08	-0.05	0.10

From Table 1 it is apparent that the bulk of the geoid signal is contained in the EGM96 model. The contribution of the residual quasi-geoid has an RMS value of only 67 cm and is, as expected, mainly of short wavelength. There is at least one noticeable inconsistency – in southeastern Ethiopia there is a high of 7 m. An initial review of the relevant data file for this region indicates that there is uncertainty in the gravity datum used and hence the results for south-eastern Ethiopia should be treated with some caution.

4.3 G_1 Contribution

This is the most computation-intensive part of the creation of the full quasi-geoid model. The first step involves the computation of a 1'x1' grid of G_1 anomalies. An interpolated 1'x1' free-air anomaly grid was used together with a 1'x1' DEM in a 2D convolution in which the horizontal distance ℓ is restricted to 40 km.

The second part of the computation involves the 2D convolution of these anomalies with the Stokes' kernel. At this point the G_1 anomaly field is re-sampled to a 5'x5' grid, so that the resultant contribution is consistent with the residual quasi-geoid. Summary statistics for the G_1 contribution to the quasi-geoid are shown in Table 1.

It is apparent from the table that the G_1 contribution is very small (RMS of 7cm). Although not shown here, a graphical representation of the G_1 contribution shows that the contribution is only significant in areas of rough terrain (the Atlas mountains in Morocco and the Ethiopean highlands are two examples). This is consistent with what has been determined previously (Amod and Merry 2002b).

4.4 Final Geoid Model

The full geoid model for Africa was created by combining the EGM96 representation, the residual quasi-geoid, the G_1 contribution and the spherical harmonic expansion of the N-ζ term (equations 1 and 2) on a 5'x5' grid, for each of southern and northern Africa. These two grids were then combined into one grid for all of Africa. This result is shown in Figure 2 and has been labelled AGP2003 (African Geoid Project 2003).

This representation shows basically the same shape as that given by the EGM96 model, with more detail, especially in the mountainous regions and in those regions for which new gravity data (not used in EGM96) were available.

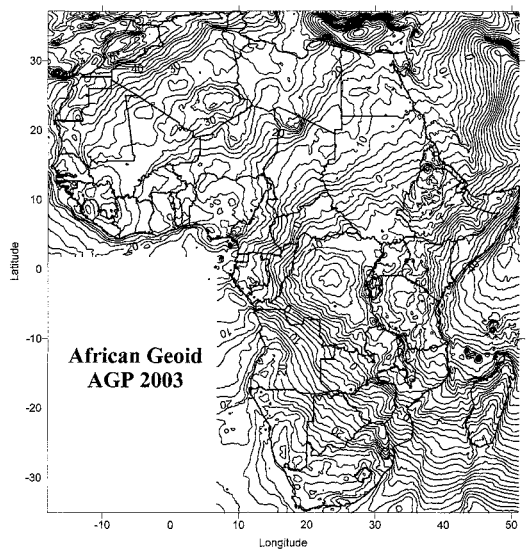


Fig. 2 Preliminary Geoid Model for Africa (contour interval: 2 metres)

4.5 Comparison with GPS/Levelling

No formal precision estimates are available for this geoid model. It is estimated that the global accuracy of the EGM96 quasi-geoid is of the order of 45 cm, with regional errors reaching 80 cm (Lemoine et al. 1998). Although our model should be an improvement on EGM96, any medium and long wavelength errors in EGM96 will be carried through into this model.

It is possible to use a combination of precise GPS heighting and levelling to evaluate the quality of the geoid. The GPS heighting provides ellipsoidal heights h and the precise levelling provides orthometric heights H. Then the geoidal height N is given by:

$$N = h - H \qquad (7)$$

GPS measurements have been carried out in many parts of Africa. However, there are very few consistent sets of GPS measurements undertaken at precise levelling benchmarks. We have been able to obtain two such sets, one covering a small part of the southwestern corner of South Africa, the other covering (rather sparsely) the eastern part of Egypt.

The AGP2003 geoid model was interpolated at some 42 GPS/levelling benchmarks in the western Cape province of South Africa, using simple bi-linear interpolation, and these values were com-

pared with the values computed at the benchmarks. The results are summarised in Table 2, together with the results obtained using EGM96 only. A similar comparison has been carried out using the eight Egyptian points, and these results are shown in Table 3.

Table 2. Differences with GPS/Levelling - western Cape Province

Model	Min. (cm)	Max. (cm)	Mean (cm)	Std. Dev. (cm)
AGP2003	- 76	- 38	- 63	9
EGM96	- 97	- 62	- 55	20

In the western Cape province of South Africa both the AGP2003 and EGM96 geoids show a bias of around 60 cm with respect to the GPS/levelling data. Although part of this bias could be due to a long wavelength error in EGM96, a significant portion of this bias is believed to be due to errors in both the GPS and the levelling datums. In the case of the GPS data, the GPS heights have been determined indirectly via the national GPS network from the ITRF VLBI station at Hartebeesthoek, near Pretoria. Recent investigations (Chandler 2001) have shown that in the western Cape the GPS-derived ellipsoidal heights are too low by about 32 cm. On the other hand, the vertical levelling datum in South Africa is approximately 20 cm below mean sea level (Merry 1990). If both these biases are removed (a combined correction of 52 cm), then the bias between the AGP2003 and the GPS/levelling data is much reduced.

The standard deviations of the differences are small (20 cm or less). This may be expected, as the 42 GPS/levelling points cover a very small area (about 150 km by 150 km), and any long wavelength errors would be absorbed by the bias. However, it is significant that the AGP2003 model reduces the standard deviation from 20 cm (for EGM96) to 9 cm. This indicates that the short wavelength features of the geoid are well-modelled by AGP2003.

Table 3. Differences with GPS/Levelling - Egypt

Model	Min. (cm)	Max. (cm)	Mean (cm)	Std. Dev. (cm)
AGP2003	- 23	+ 199	+ 124	80
EGM96	- 160	+ 118	+ 5	97

The eight Egyptian GPS/levelling points are distributed throughout the eastern part of the country, and the discrepancies are much larger than for the small region in South Africa. As with the South African data, there are potentially errors in both the GPS and levelling data. However, long wavelength errors in EGM96 will also play a significant role. An analysis of the difference between GPS/levelling and EGM96 using additional data (Abd-Elmotaal 2002) shows that there are significant long wavelength discrepancies. Although AGP2003 shows a large bias with respect to the GPS/levelling data, it is somewhat more consistent than EGM96.

5. Conclusions and Recommendations

Gravity data for Africa have been successfully combined with a global geopotential model in a remove-restore procedure to determine a representation of the geoid for the whole of Africa - AGP2003. Two-dimensional convolution with a spherical Stokes' kernel has been used to evaluate the contribution of residual gravity anomalies. The Molodensky G_1 term and its contribution have also been calculated, using 1'x1' gravity anomaly and height grids. Finally the geoid/quasi-geoid separation has been restored using Rapp's harmonic representation of this correction.

Comparison of the results with GPS/levelling data covering Egypt and a small part of South Africa shows that the AGP2003 model is a significant improvement over the EGM96 global model, especially in South Africa. The standard deviation of the difference is reduced from 20 cm to 9 cm in South Africa and from 97 cm to 80 cm in Egypt.

The biggest weakness in this model is the lack of data for a large part of the continent. A significant improvement in the quality of the model will only be realised once these gaps have been filled.

Acknowledgements

The Getech and UCT gravity data sets are compilations, derived from a large number of sources, too many for all of them to be listed here. The marine gravity data were provided by the Danish Kort & Matrikelstyrelsen while the GLOBE DEM was provided by the NOAA National Geophysical Data Center.

Financial support for this research was provided by the University of Cape Town.

References

Abd-Elmotaal, H. (2002). Towards a precise geoid for Egypt. Presented, Third meeting of the Inter-

national Gravity and Geoid Commission, Thessaloniki, Greece, August 2002.

Amod, A. and C.L. Merry (2002a). The use of the two-dimensional spherical FFT for quasi-geoid modelling in South Africa. *Survey Review* **36**(285):508-520.

Amod, A and C.L.Merry (2002b). A note on the Molodensky G_1 term. *Bulletin of the International Geoid Service* **12**: 52-80.

Andersen, O.B. and P. Knudsen (1998). Global marine gravity field data from the ERS-1 and Geosat geodetic mission altimetry. *Journal of Geophysical Research,* **103**(C4): 8129.

Benahmed Daho, S.A. and S. Kahlouche (2000). The new gravimetric geoid in Algeria. *Bulletin of the International Geoid Service* **10**: 59-77.

Berry, P.A.M., J.E. Hoogerboord, R.A. Pinnock (1999). An independent global assessment of the quality of current global digital elevation models. Paper presented at the XXII General Assembly of the International Union of Geodesy and Geophysics, Birmingham, United Kingdom, July 1999.

Chandler, G. (2001). Accuracy assessment of TrigNET base station co-ordinates. BSc thesis, Department of Geomatics, Univ. of Cape Town.

Fairhead, J.D., A.B. Watts, P. Chevalier, B. El-Haddadeh,, C.M. Green, G.W. Stuart, K.A. Whaler, I. Whindle (1988). African Gravity Project. Technical Report, University of Leeds Industrial Services Ltd., Leeds, United Kingdom.

Fashir, H.H., A. Majid, A. Kadir (1998). African regional gravimetric geoid 1997: Stokesian approach. *Bulletin of the International Geoid Service* **8**: 65-84.

Featherstone, W.E., J.F. Kirby, A.H.W. Kearsley, J.R. Gilliland, G.M. Johnston, J. Steed, R. Forsberg, M.G. Sideris (2001). The AUSGeoid98 geoid model of Australia: data treatment, computations and comparisons with GPS-levelling data. *Journal of Geodesy* **75**:313-330.

Forsberg, R.and C.C. Tscherning (1981). The use of height data in gravity field approximation. *Journal of Geophysical Research* **86**:7843-7854.

Gachari, M.K. and J.G. Olliver (1998). A high resolution gravimetric geoid of the eastern Africa region. *Survey Review* **34**(269):421-436.

Hastings, D.A. and P.K. Dunbar (1998). Development and assessment of the Global Land One-km Base Elevation digital elevation model (GLOBE). *ISPRS Archives* **32**(4):218-221.

Heiskanen, W.A. and H. Moritz (1967). *Physical Geodesy.* W.H. Freeman, San Francisco.

Lemoine, F.G., S.C. Kenyon, J.K. Factor, R.G. Trimmer, N.K. Pavlis, D.S. Chinn, C.M. Cox, S.M. Klosko, S.B. Luthcke, M.H. Torrence, Y.M. Wang, R.G. Williamson, E.C. Pavlis, R.H. Rapp, T.R. Olsen (1998). The development of the joint NASA GSFC and NIMA geopotential model EGM96. NASA Technical Publication NASA/TP-1998-206861. Goddard Space Flight Center, Greenbelt, MD, USA.

Merry, C.L. and H. van Gysen (1987). A regional geoid for southern Africa. Presented, XIX General Assembly of the International Union of Geodesy and Geophysics, Vancouver, Canada.

Merry, C.L. (1990). Recent variations in Mean Sea Level in Southern Africa. In: *Sea Surface Topography and the Geoid* (Eds: H Sünkel & T Baker), Springer-Verlag, New York:149-157.

Merry, C.L. and D. Blitzkow (2001). The African Geoid Project. Presented, Scientific Assembly of the International Association of Geodesy, Budapest, Hungary.

Merry, C.L. (2003). DEM-Induced Errors in Quasi-Geoid Modelling in Africa. Accepted, *Journal of Geodesy.*

Obenson, G. (1985). Geoidal heights for Africa from GEM9 potential coefficients and degree gravity anomalies. *Survey Review* **28**(218):207-221.

Rapp, R.H. (1997). Use of potential coefficient models for geoid undulation determination using a spherical harmonic representation of the height anomaly/geoid undulations difference. *Journal of Geodesy*, **71**(5):282-299.

Roman, D.R. and D.A. Smith DA (2000). Recent investigations towards achieving a one centimetre geoid. *Proceedings, GGG2000*, Banff, Canada.

Strang van Hees, G. (1990). Stokes formula using fast Fourier techniques. *Manuscripta Geodaetica,* **15**(4):235-239.

van Gysen, H.G. and C.L. Merry (1987). The gravity field in southern Africa. Univ. of Cape Town Dept. of Surveying Technical Report TR-3.

Optimal Marine Geoid Determination in the Atlantic Coastal Region of Argentina

C. Tocho
Facultad de Ciencias Astronómicas y Geofísicas, Paseo del Bosque s/n, 1900 La Plata, Argentina.

G.S. Vergos
Department of Geodesy and Surveying, Aristotle University of Thessaloniki, Univ. Box. 440, 54124, Thessaloniki, Greece.

M.G. Sideris
Department of Geomatics Engineering, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, T2N 1N4, Canada:

Abstract. The determination of an optimal, in terms of resolution and accuracy, marine geoid model for the Atlantic coastal region of Argentina is investigated using satellite altimetry and shipborne gravity data. The altimetric data used are those of the geodetic phase of the ERS1 mission while marine gravity data have been employed as well to determine a gravimetric geoid solution. Furthermore, the effect of the Quasi-Stationary Sea Surface Topography (QSST) was taken into account in correcting the altimetric Sea Surface Heights (SSHs) to derive geoid undulations. Special emphasis was placed on reducing the effects of the Sea Surface Variability (SSV) on the densely spaced altimetric SSHs with low-pass filtering. The satellite and shipborne data were combined in the spectral domain to improve the accuracy of the altimetric solution close to the coastline and derive a more rigorous solution. The accuracy of the final geoid models is assessed through comparisons with stacked TOPEX/POSEIDON (T/P) SSHs, known for their high precision. From the results achieved it was concluded that an altimetric geoid accurate to about 5-8 cm (1σ) is feasible in some areas, while the gravimetric solution gives poorer results by about 5-6 cm. The combination of satellite and shipborne data with the proposed algorithm improves the accuracy of the gravimetric geoid model by about 2 cm.

Keywords. Marine geoid, altimetry, Sea Surface Heights, shipborne gravity, MIMOST.

1 Introduction

The main objective of this paper is the determination of a high-accuracy and high-resolution marine geoid model in the Atlantic coastal region of Argentina. The theoretical background related to the estimation of the gravimetric and altimetric geoid models (Vergos 2002) and the combined one using the Multiple Input Multiple Output System Theory (MIMOST) (Sideris 1996, Andritsanos and Tziavos 2002), will be outlined together with the description of the data available in the area under study. Finally, some numerical studies carried out in this research will be presented.

2 Computational methodology

2.1 Gravimetric geoid modeling

The gravimetric geoid determination was based on shipborne and satellite altimetry-derived gravity anomalies. The latter have been used to augment the ship data and fill-in gaps. The shipborne gravity data referred to the Geodetic Reference System 1967 (GRS67), thus they had to be transformed to GRS80. This was done using the following basic formulae (Moritz 2000):

$$\Delta g_{GRS80} = \Delta g_{GRS67} + \gamma_{GRS67} - \gamma_{GRS80} \quad (1)$$

where, Δg_{GRS80} denotes gravity anomaly in GRS80, Δg_{GRS67} gravity anomaly in GRS67 and γ_{GRS67} and γ_{GRS80} are the magnitudes of the normal gravity in GRS67 and GRS80 respectively. Normal gravity can be computed for GRS67 and GRS80 as (*ibid.*)

$$\gamma_{GRS80} = 978032.6702 \frac{1 + 0.001931851 \sin^2 \varphi}{\sqrt{1 - 0.00669438002 \sin^2 \varphi}} \quad (2)$$

It is often that the ship gravity anomalies refer to the sea surface and not the geoid itself, thus their use will lead to the determination of a mean sea surface and not a geoid model. These gravity anomalies have to be free-air reduced so as to produce gravity anomalies on the surface of the geoid. The free-air gravity anomalies are computed using the well-known reduction formula

$$\Delta g_f = \Delta g - \delta g_f \quad (3)$$

where δg_f is the free-air reduction. In marine regions the height needed for the reduction is that of the QSST, which varies between 0 and -0.30 m in the study area. Thus, for practical purposes it is sufficient to use the normal gradient of gravity to compute the free-air reduction as

$$\delta g_f \approx -\frac{\delta\gamma}{\delta h} h \approx -0.3086h \quad (4)$$

where h is the QSST in meters, derived from a global model.

These two pre-processing steps are necessary for the data homogenization so that they can be used for the determination of the gravimetric geoid. The marine gravimetric geoid will be computed using the remove-compute-restore technique employing Stokes' formula for the prediction of residual geoid heights. Before the prediction of the geoid the gravity anomalies have to be reduced to a geopotential model during the remove step. Furthermore the effect of the topography, actually that of the bathymetry in marine areas, has to be taken into account through a topographic reduction. In this study, a residual terrain model (RTM) reduction was used to account for the bathymetry. The RTM effect on gravity is given by the approximate expression (Forsberg 1984)

$$\Delta g_{RTM} \approx 2\pi G\rho(h - h_{ref}) - c \quad (5)$$

where h is the bathymetric depth given by a global bathymetry model, h_{ref} is the depth of a smooth mean reference surface and ρ is the density contrast between Earth's crust and seawater. The reference bathymetric surface was obtained by simple averaging the fine bathymetry grid and then low-pass filtering it using a moving-average window, with a resolution around 100 km. The residual gravity anomalies are then gridded and the contribution of the bathymetry is restored prior to the calculation of the geoid height. The bathymetry refers to masses below the geoid so its effect has to be restored before the use Stokes formula for the estimation of geoid heights (Dahl and Forsberg 1998).

Different approximations to Stokes' kernel function were investigated to compute residual geoid undulations all in the spectral domain. Finally, it was decided to employ the 2D-FFT spherical Stokes convolution to evaluate the kernel function (Strang Van Hess 1990)

$$N^{gr} = \frac{R\Delta\varphi\Delta\lambda}{4\pi\gamma} \mathcal{F}_1^{-1}\left\{\{\mathcal{F}\{\Delta g\cos\varphi\}\mathcal{F}(\Delta\varphi,\Delta\lambda,\varphi_m)\}\right\} \quad (6)$$

where N^{gr} are the estimated residual gravimetric geoid heights and $\mathcal{F}$, $\mathcal{F}^{-1}$ denote the direct and inverse Fourier transforms. The final gravimetric geoid is obtained by restoring both the contribution of the reference model and the RTM effects to geoid heights.

2.2 Altimetric geoid modeling

An altimetric satellite measures the time taken by a radar pulse to travel from the satellite to the sea surface and then back to the satellite receiver. Combined with precise satellite location data, altimetry measurements yield Sea Surface Heights (SSHs). The so-derived SSHs have to be corrected for several geophysical effects (tides, tidal loading, ionosphere, wet and dry troposphere, inverse barometer and electromagnetic bias) and instrumental errors (ultra-stable oscillator, centre of gravity, corrections for instrument and algorithm effects that can not be modeled and waveforms). After applying the above corrections, Corrected Sea Surface Heights (CORSSHs) are available for one or more satellites (e.g. GEOSAT, ERS1).

Sea Surface Heights contain information about both the geoid and the sea surface topography (SST), while the latter consists of a time-dependent and a nearly time-independent component (quasi-stationary part). Stacking the repeat tracks can eliminate the effect of the time-dependent component and part of the sea surface variability effects that influence the data.

These altimetric measurements refer to the sea surface so they have to be reduced to the geoid. This is performed by estimating the QSST at each sub-satellite point and removing the contribution of the QSST from the SSH value. The quasi-stationary component of the SST is modeled by a spherical harmonic series of the Dynamic Ocean Topography (DOT) as follows:

$$\varsigma_c(\phi,\lambda) = \left[\sum_{n=1}^{n_{max}}\sum_{m=0}^{n}\left(\bar{C}_{nm}^{*SST}\cos m\lambda + \bar{S}_{nm}^{SST}\sin m\lambda\right)\bar{P}_{nm}(\sin\phi)\right] \quad (7)$$

where $\varsigma_c(\phi,\lambda)$ is the contribution of the model coefficients, n_{max} denotes the maximum degree and order of expansion of the DOT model, $\bar{P}_{nm}(\sin\phi)$ are the fully normalized associated Legendre functions, and $\bar{C}_{nm}^{*SST}$, $\bar{S}_{nm}^{SST}$ are the fully normalized DOT spherical harmonic coefficients.

After the removal of the effect of the QSST the SSHs refer to the geoid and can be used to derive an altimetric geoid model. As in the gravimetric geoid computation, the contribution of a geopotential model was removed to derive reduced SSHs

($SSHs_{red}$). The so-reduced SSHs ($SSHs_{red}$) may still contain some blunders, so a 3 rms test is used to identify and remove gross-errors. If the mean value of the reduced SSHs is small enough (e.g., bellow 10 cm), then the 3 rms test can be applied. That is so because by using a 3 rms test we assume that all systematic errors have been removed from the data and only random errors remain. If the mean value of the SSH_{red} is larger, an RTM reduction is applied first to obtain smoother residual SSHs. The computation of the RTM effects on residual geoid heights has been based on the same concept as in the gravimetric geoid. In both cases the GRAVSOFT software (Tscherning et al. 1992) has been used to create the reference bathymetric grid and estimate the RTM reduction on gravity anomalies and geoid heights.

The residual Sea Surface Heights represent the medium wavelengths of the geoid heights and can be considered as residual geoid heights (N_{res}). After all these processing steps the N_{res} are ready to be gridded. The gridding was based on a weighted means method using the inverse of the square of the distance as the weight for each irregular observation. If the area under study is located in a closed sea, then the such-derived N_{res} constitute the final estimated residual altimetric geoid height. But, if in an open-ocean area, then the effect of the Sea Surface Variability (SSV) influences the data and appears as noise in the N_{res}. Such effects need to be removed or at least reduced so as to derive reliable predictions. Since the effect of the SSV appears as high-frequency noise in the altimetric data, it can be reduced by low-pass filtering the grid of N_{res}. This was performed using a Wiener filtering, which is equivalent to least squares collocation. Then, the final step to compute the altimetric geoid N^{alt} is to restore the contribution of the geopotential model and the contribution of the bathymetry.

2.3 Combined geoid modeling

Using the Multiple Input Multiple Output System Theory (MIMOST) for the optimal combination of heterogeneous noisy data a combined solution has been determined. That was done in order to investigate whether the combined use of shipborne gravity and satellite altimetry data improves the geoid compared to the purely gravimetric case. The algorithm and the related formulae are given in Sideris (1996) and Andritsanos and Tziavos (2002).

The input signals in the combined solution were the residual gravimetric and altimetric geoid heights before the restore of the contribution of the geopotential model. That was done to avoid introducing long-wavelength errors. Since no information was available about the errors of the input data, simulated noise was used as input errors in the prediction. White noise was generated in MATLAB® using as standard deviation (σ) that of the differences of each solution with T/P SSHs.

3 Area under study and data availability

The area under study was located in the Atlantic Coastal region of Argentina, bounded by 34°S to 55°S in latitude and 56°W (304°E) to 70°W (290° E) in longitude. The marine gravity data available were 17352 gravity anomalies provided by BGI (Bureau Gravimétrique International). Since there were some gaps between the ship tracks, the KMS01 2′×2′ altimetry derived free-air gravity anomaly field (Andersen and Knudsen 1998) has been used as fill-in information. The distribution of the marine gravity data is depicted in Figure 1. KMS01 is the newest compilation of a global altimetry-derived marine free-air gravity field by the KMS group at the Danish Surveying and Cadastre.

To derive the long-wavelength information of the gravity field we used the EGM96 global geopotential model (Lemoine et al. 1998) complete to degree and order 360. The topographic/bathymetric data for the RTM reduction were those of the Smith and Sandwell model (Smith and Sandwell 1997), which resulted as a combination of depths derived from altimetry.

The QSST was computed from the EGM96 DOT model, which is a spherical harmonic expansion of the SST, complete to degree and order 20 (Lemoine et al. 1998). That model was derived during the simultaneous adjustment for the development of the EGM96 geopotential model and its representation of the QSST for the area under study is given in Figure 2. Finally, the altimetric data were 70510 SSHs measurements from the geodetic mission (GM) of the ERS1 satellite and 80864 repeated T/P SSHs from the entire 3rd Year of the satellite's mission (Aviso 1998).

4 Geoid model development

4.1 Altimetric geoid model with ERS1-GM data

The ERS1 satellite altimetry SSHs were provided in the usual Geophysical Data Records (GDRs) format and were corrected for all geophysical and instrumental errors in a pre-processing step according to the models and methods described in the AVISO handbook (AVISO 1998).

As the ERS1-GM CORSSHs refer to sea surface, they were reduced to the geoid by removing the effect of the QSST. That was performed using Eq. 8 to predict QSST values on the irregular ERS1 points.

At this point, the ERS1-GM Sea Surface Heights referred to the geoid were ready to be used for the estimation of a purely altimetric geoid. The contribution of the EGM96 geopotential model was then removed and the resulting reduced SSHs were checked for their mean value. After examining the mean value of the reduced field it was not found to be small enough for a 3rms test to be performed. Thus, the bathymetry was first taken into account with an RTM reduction and after that a 3 rms test for blunder detection has been applied. After the 3 rms test 678 points were removed and the resulting point data were gridded using the aforementioned algorithm on a 3′×3′ grid.

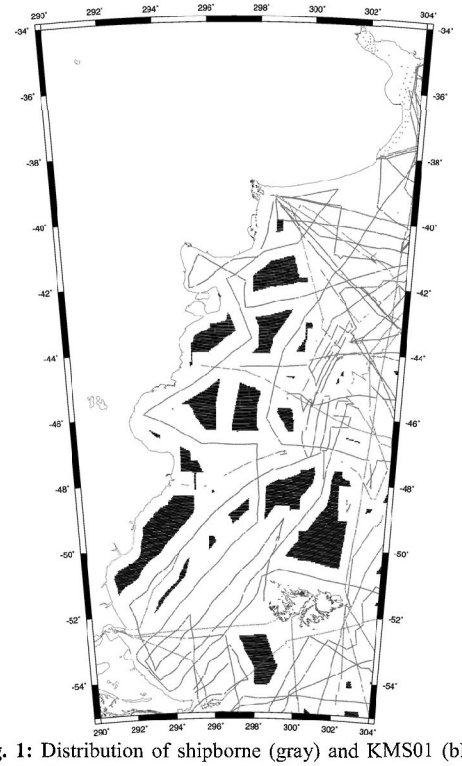


Fig. 1: Distribution of shipborne (gray) and KMS01 (black) data in the area under study.

To reduce the high-frequency SSV-like effects the data were low-pass filtered using Wiener filtering. The cut-off frequency was determined empirically based on a criterion of maximum noise reduction with minimum signal loss. A number of cut-off frequencies were tested and finally a cut-off frequency corresponding to a wavelength of 20 km was chosen. That selection gave the best results as far as both the noise reduction and the minimization of the differences with T/P SSHs are concerned.

The final altimetric geoid solution was obtained by restoring the contribution of the EGM96 geopotential model and that of the RTM effects of the bathymetry. Table 1 presents the statistics of the final ERS1-GM geoid for the area under study which is also depicted in Figure 3 for visualization purposes.

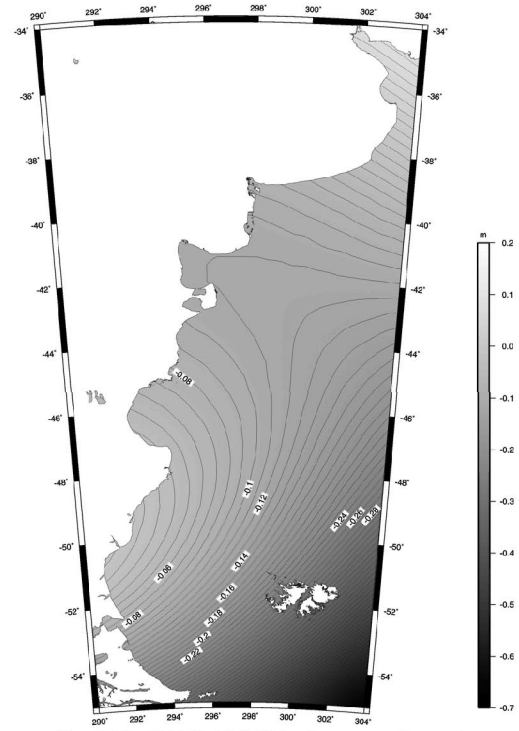


Fig. 2 The EGM96 QSST in the area under study.

Table 1. Statistics of the ERS1-GM geoid. Unit: [m].

	max	min	mean	σ
N^{alt}	19.665	0.505	11.460	±2.996

4.2 Gravimetric geoid model

The gravimetric geoid model was determined using the computational procedure described in paragraph 2.1. The main difference with the altimetric geoid modeling is that the RTM reduction is restored before the prediction of the gravimetric residual geoid heights. Table 2 presents the statistics of the final gravimetric geoid for the area under study which is also depicted in Figure 4.

Table 2. Statistics of the gravimetric geoid. Unit: [m].

	max	min	mean	σ
N^{grav}	19.333	0.642	11.362	±2.957

4.3 Combined solution

The estimation of the combined solution was performed in a smaller area between 40°S to 50°S in

latitude and 56°W (304°E) to 66°W (294° E) in longitude. Table 3 shows the statistics of the gravimetric geoid, the ERS1-GM geoid and the combined solution in the reduced area and Figure 5 shows the final solution from MIMOST. The input noise for each dataset was generated using the σ of the differences T/P SSHs and the gravimetric (±25 cm) and altimetric geoid (±20 cm) geoid models.

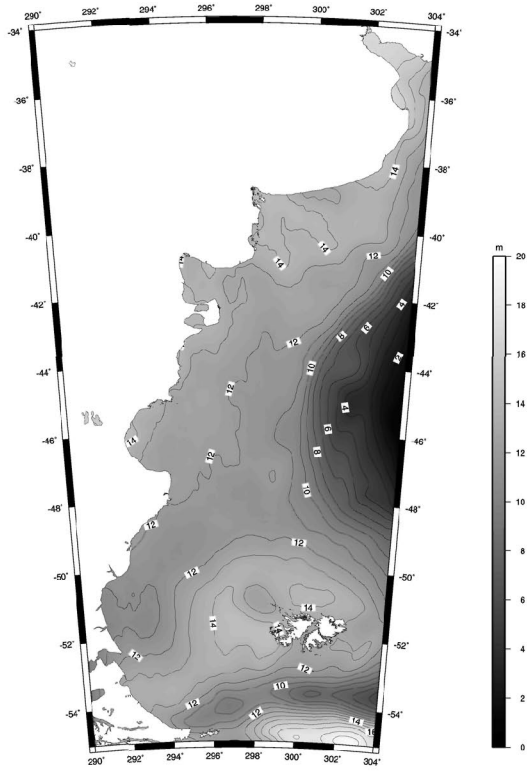


Fig 3: The final ERS1-GM altimetric geoid model for Argentina.

Table 3. Statistics of the geoid models in the inner area. Unit: [m].

	max	min	mean	σ
N^{alt}	17.225	0.642	10.253	±3.273
N^{grsv}	16.542	0.545	10.293	±3.290
N^{comb}	17.103	0.574	10.266	±3.279

5 Validation of the estimated geoid models

The accuracy of the models was assessed through comparisons with stacked T/P SSHs. The computed differences between T/P and each geoid solution were minimized using a four-parameter transformation model, as

$$N^y = N^i - b_0 \cos\varphi\cos\lambda - b_1 \cos\varphi\sin\lambda - b_2 \sin\varphi - b_3 \quad (8)$$

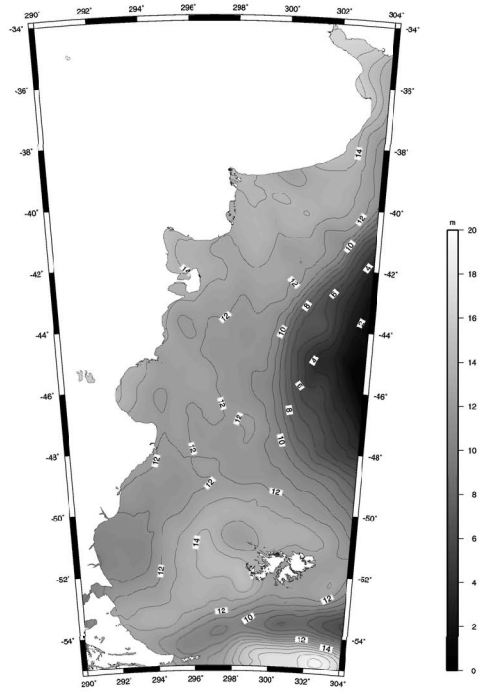


Fig 4: The final gravimetric geoid model for Argentina.

where N^y is the stacked T/P SSHs, N^i is the altimetric, gravimetric or combined geoid solution and the parameters b_0 b_1 b_2 and b_3 were calculated using a least squares technique. The statistics of the differences after the bias and tilt fit between the T/P SSHs and the estimated geoid solutions are given in Table 4. From that Table it can be seen that the overall best agreement is achieved for the altimetric geoid solution while the data combination improves the gravimetric geoid by about 2 cm.

Worth mentioning though, is that the σ of the differences for the comparisons with the altimetric models is quite high, at the ±20 cm level, while a value close to ±9 cm would be expected based on previous studies (Li and Sideris 1997; Vergos 2002). Plotting the differences it was noticed that their largest and smallest values are located close to the coastline and more specifically between $-45^\circ \le \varphi \le -44^\circ$ and $294^\circ \le \lambda \le 296^\circ$ and close to Falkland Islands where the effects of SSV and other oceanic phenomena are very strong. In the rest of the region, the differences are within their expected values ranging between -60 and 60 cm. In our opinion, this is an indication that the accuracy of the altimetric geoid models is much better than the comparisons with T/P imply. Neglecting a few T/P points that refer to the aforementioned regions the σ of the differences reduces to about ±5 to ±8 cm for the altimetric geoid models. The same improvement of more than ±9 cm holds for the gravimetric and combined models too.

So, it can be concluded that by only stacking the T/P data part of the oceanic effects, which clearly influence the SSHs used for the comparisons, cannot be removed. Probably, the T/P data had to be low-pass filtered as well in their along-track direction, to further reduce the effect of the SSV and make the comparisons more representative.

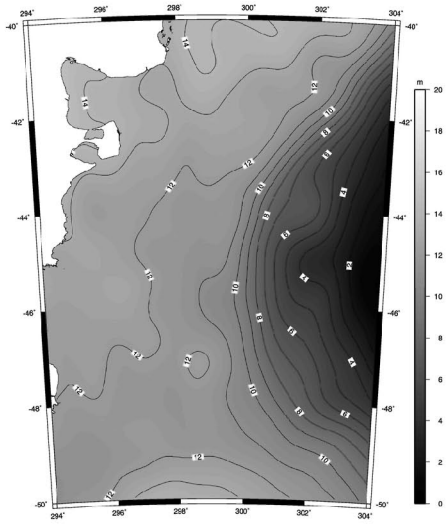


Fig 5: The combined geoid model.

Table 4. Geoid height difference between T/P and the estimated models. Unit: [m].

	max	min	mean	σ
$N^{T/P} - N^{ERS1}$	1.00	-1.20	0.00	±0.20
$N^{T/P} - N^{gr}$	1.66	-0.93	0.00	±0.25
$N^{T/P} - N^{comb}$	1.49	-0.93	0.00	±0.23

6 Conclusions

Altimetric, gravimetric and a combined final geoid height solution have been determined in the Atlantic coastal region of Argentina. The MIMOST theory for the optimal combination of heterogeneous data has been applied to improve the gravimetric geoid solution close to the coastline.

From the results and validation procedures carried out, it is evident that when altimetry and shipborne gravity data are handled properly, i.e. corrected for all error sources, blunders removed, accurate geopotential and DOT models used, the data are corrected for the QSST signal, the bathymetry is taken into account using an accurate model, the altimetry data are low-pass filtered, then, altimetric geoid modeling accurate to about ±7 cm is feasible, while the combined solution improves the gravimetric one, by about 2 cm, in terms of the σ of the differences with T/P SSHs. These differences refer to purely oceanic areas (not close to the coastline) and regions where the effect of the variability of the oceans is not very strong. Some new results must be acquired using GEOSAT-GM data.

The effect of oceanic phenomena in the densely spaced GM datasets, especially in areas with high ocean dynamics, is profound and should be reduced by low-pass filtering the altimetric datasets. If this step is neglected, then the resulting geoid solutions are less accurate by about 2-5 cm. A point that needs further research is on the use of crossover adjustment for the reduction of such noisy signals. Finally, the altimetry data should be corrected for the QSST signal to refer to the geoid and not the sea surface, while the question that arises is not on the necessity of such a correction, but on the selection and the development of accurate DOT models.

References

Andersen OB, Knudsen P (1998) Global gravity field from ERS1 and Geosat geodetic mission altimetry. J Geophys Res 103(C4): 8129-8137.

Andritsanos VD, Tziavos IN (2002) Estimation of gravity field parameters by a multiple input/output system. Phys and Chem of the Earth, Part A 25(1): 39-46.

AVISO User Handbook (1998). Corrected Sea Surface Heights (CORSSHs). AVI-NT-011-311-CN, Edition 3.1.

Dahl OC, Forsberg R (1998) Geoid models around Sognefjord using depth data. J Geod 72: 547-556.

Forsberg R (1984) A study of terrain corrections, density anomalies and geophysical inversion methods in gravity field modeling. Report of the Dept. of geodetic Science and Surveying No 355. The Ohio State Univ., Columbus, Ohio.

Lemoine, F. G., Kenyon, S. C., Factim, J. K., Trimmer, R. G., Pavlis, N. K., Chinn, D. S., Cox, C. M., Klosko, S. M., Luthcke, S. B., Torrence, M. H., Wang, Y. M., Williamson, R. G., Pavlis, E. C., Rapp, H. and Olson, T. R., 1998. The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM 96. Pub. Goddard Space Flight Center.

Li J, Sideris MG (1997) Marine gravity and geoid determination by optimal combination of satellite altimetry and shipborne gravimetry data. J Geod 71(1): 209-216.

Moritz H (2000) Geodetic Reference System 1980. J Geod 74(1): 128-162.

Sideris MG (1996) On the use of heterogeneous noisy data in spectral gravity field modeling methods. J Geod 70(8): 470-479.

Smith WHF, Sandwell DT (1997) Global Sea Floor Topography from Satellite Altimetry and Ship Depth Soundings. Science Magazine, vol. 277, Iss 5334.

Strang van Hees G (1990). Stokes' formula using fast Fourier techniques. Manuscripta Geodaetica, Vol. 15, pp. 235-239.

Tscherning CC, Forsberg R, Knudsen P (1992) The GRAVSOFT package for geoid determination. In: Holota P, Vermeer M (eds) 1st Continental Workshop on the Geoid in Europe, Prague, June 7-9, 1993, pp 327-334.

Vergos GS (2002) Sea Surface Topography, Bathymetry and Marine Gravity Field Modelling. MSc Theses Dissertation, Dept of Geomatics Engineering, University of Calgary, UCGE Reports 20157, Calgary, Alberta.

Comparison of Newton's Integral in the Space and Frequency Domains

M. Kuhn[1], K. Seitz[2]
[1]Western Australian Centre for Geodesy, Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia, e-mail: kuhnm@vesta.curtin.edu.au, Fax: +61-8-9266-2703
[2]Geodetic Institute, University of Karlsruhe, Englerstrasse 7, PO Box 6980, D-76128 Karlsruhe, Germany, e-mail: seitz@uni-karlsruhe.de, Fax: +49-721-6086552

Abstract. Newton's integral can be evaluated either in the space domain by a direct integration approach or in the frequency domain by a spherical harmonic approach. In the first approach Newton's integral is evaluated by a *discretised* numerical integration using the gravitational potential (or its derivatives) caused by regularly shaped bodies such as prisms or tesseroids (i.e., spherical or elliptical volume elements, respectively). The second approach expresses Newton's integral by spherical harmonic expansions of height or density functions, which can be interpreted as the gravitational effect of different mass layers. This paper studies the theoretical and practical differences between both approaches. The gravitational effect of the global topographic masses, described here by equivalent rock heights, is evaluated using both approaches and the results compared. Numerical results for the gravitational potential are given for evaluation points located directly on the Earth's surface and at different heights (e.g., flight levels) above it.

Keywords. Newton's integral, numerical integration, spherical harmonics, tesseroid, topography

1 Introduction

The computation of gravitational effects (e.g., potential or gravity) caused by different mass distributions is a central part in the study of the Earth's gravity field. (e.g., Heiskanen and Moritz 1967, Lambeck 1988). The gravitational potential is given by Newton's (volume) integral, which in practice can be evaluated in many different ways. Here a classification will be made between space-domain and frequency-domain techniques.

In the space-domain, Newton's integral is solved either analytically (in the case that a closed solution exists for the regarded mass distribution) or numerically. Analytical formulae can be given, e.g., for simple geometrical bodies such as right-rectangular prisms (e.g., Mader 1951, Nagy 1966, Nagy et al. 2000 and 2002). However, in most cases the regarded mass distribution has a more complex structure (e.g., Earth's mass distribution), which makes an analytical solution of Newton's integral impossible. In this case, numerical integration can be used.

The numerical integration can be performed either by applying a standard algorithm based on the Gaussian quadrature rule (e.g. Hildebrand 1974) or dividing the regarded mass distribution into a fixed number of regular-shaped bodies of which the gravitational effect can be evaluated analytically. In the latter case, Newton's integral is evaluated by summing up the effects of all bodies forming the regarded mass distribution. This approach will be called here *discretised* numerical integration.

A different way to evaluate Newton's integral is given by its transformation in the frequency domain. In this approach, Newton's integral is expressed by spherical harmonic expansions of height or (surface) density functions, which can be interpreted as the gravitational effect of different mass layers (Kuhn and Featherstone 2003a). This approach is well suited to express the gravitational potential caused by global mass distributions defined by their extension above/below (height/depth) a given mean sphere and a 2- or 3-D mass density distribution (e.g., Ramillen 2002, Kuhn and Featherstone 2003b, this issue).

In this paper, the principal differences between both the *discretised* numerical integration and spherical harmonic approach will be studied using the gravitational effect caused by the global topographic masses defined by the equivalent rock heights of the 5-arc-min x 5-arc-min digital elevation model (DEM) JGP95E (Lemoine et al. 1998). Numerical results are given for the gravitational potential, derived by the *discretised* numerical integration approach using tesseroids (i.e. spherical or elliptical volume elements) and by the spherical harmonic approach at globally distributed, co-located evaluation points.

2 Methodology

2.1 Discretised Numerical Integration of Newton's Integral

Newton's integral is evaluated by *discretised* numerical integration dividing the topographic masses into spherical volume elements. In a spherical approximation, this is a somewhat natural way to express the topographic masses (or other mass distributions), which usually are given by a DEM with respect to a geographical coordinate system. The spherical volume element, illustrated in Fig. 1, is bounded by two meridional planes corresponding to the spherical latitudes λ_1 and λ_2, two coaxial cones, defined by the spherical latitudes φ_1 and φ_2 and two concentric spheres defined by the radii $r_1 = R$ and $r_2 = r_1+H$, where R stands for the mean Earth radius. For such a geometrical body, the generic term "tesseroid", introduced by Anderson (1976), will be used.

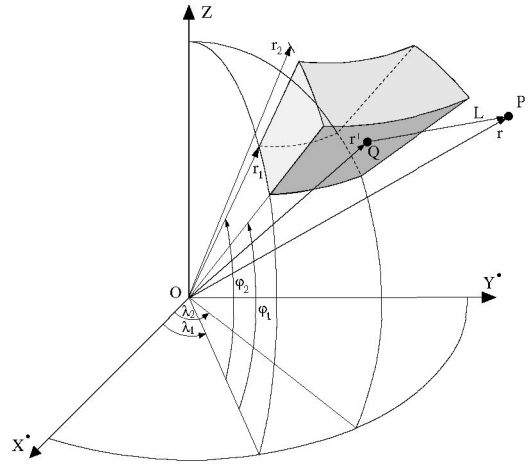


Fig. 1: Geometry of a spherical tesseroid

The gravitational potential caused by a spherical tesseroid can be derived from Newton's volume integral

$$V(\Omega',r) = G \iiint_{v_{Tess}} \frac{\rho(\Omega')}{l}\, dv \qquad (1)$$

where $dv = r^2\cos\varphi d\lambda d\varphi dr$ is the spherical volume element, r stands for the radial distance measured from the Earth's centre to the point $P(\Omega',r)$ and Ω' stands for the coordinate pair (λ,φ). In Eq. (1) $\rho(\Omega')$ denotes a lateral variable density function, which is constant within each tesseroid. In order to evaluate Newton's integral [Eq. (1)], a series expansion of the integrand around the mid-point $P_0=(\lambda_0=(\lambda_1+\lambda_2)/2,\ \varphi_0=(\varphi_1+\varphi_2)/2,\ r_0=(r_1+r_2)/2)$ is performed (e.g. Seitz and Heck 2003). A piece-wise integration yields

$$V(\Omega',r) = G\rho(\Omega')r_0 \frac{\cos\psi_0}{\sqrt{C}} \Delta\lambda\Delta\varphi\Delta r + \quad O^2(\Delta\lambda,\Delta\varphi,\Delta r) \qquad (2)$$

with the constant $C = l_0^2 / r_0^2$ and $l_0 = 2\pi r\sin(\psi_0/2)$, where ψ_0 is the spherical distance between the mid-point P_0 of the tesseroid and computation point $P(\Omega',r)$ (e.g. Heiskanen and Moritz 1967). A more detailed discussion of this approach including the formulae up to third-order will be given in a further publication.

2.2 Newton's Integral in spherical harmonics

A different approach to evaluate Newton's integral is given by its transformation in the spectral domain. In a spherical approximation, this is given by a spherical harmonic expansion of Newton's integral. A detailed description of this approach can be found e.g. in Ramillen (2002) or Kuhn and Featherstone (2003a). Only the principles will be given here.

In this approach, the gravitational potential $V(\Omega,r)$ caused by a given mass-density distribution at point $P(\Omega,r)$ located on or above the mean sphere with radius R is expressed by

$$V(\Omega,r) = \sum_{n=0}^{N_{max}} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^{n} \overline{V}_{nm}\overline{Y}_{nm}(\Omega) \qquad (3)$$

with the fully normalised spherical harmonic coefficients $\overline{V}_{nm}$ and the surface spherical harmonics

$$\overline{Y}_{nm}(\Omega) = \begin{cases} \overline{P}_{nm}(\cos\theta)\cos m\lambda & m \geq 0 \\ \overline{P}_{nlml}(\cos\theta)\sin|m|\lambda & m < 0 \end{cases} \qquad (4)$$

for the associated Legendre functions of the first kind $\overline{P}_{nm}(\cos\theta)$ for degree n and order m (e.g., Heiskanen and Moritz 1967). Here Ω stands for the coordinate pair (λ,θ) with the co-latitude (polar distance) θ. The fully normalized spherical harmonic coefficients are given by an infinite sum of spherical harmonic coefficients $\overline{\kappa}_{p,nm}$ representing surface density functions (e.g., Ramillen 2002, Kuhn and Featherstone 2003a)

$$\overline{V}_{nm} = \mu_n \left(\frac{R_L}{R}\right)^{n+2} \sum_{p=1}^{\infty} \frac{\Gamma(n+3)}{p!\,\Gamma(n+4-p)} \overline{\kappa}_{p,nm} \qquad (5)$$

with the factor

$$\mu_n = \frac{4\pi GR}{2n+1} \qquad (6)$$

the gamma function $\Gamma(\cdot)$ and the fully normalized spherical harmonic coefficients $\bar{\kappa}_{p,nm}$ of the surface density functions

$$\kappa_p(\Omega) = \rho(\Omega)\frac{H(\Omega)^p}{R_L{}^{p-1}}, \qquad p \in N^+. \tag{7}$$

In Eq. (5) the actual mass distribution used has been described by the lateral variable functions $\rho(\Omega)$ and $H(\Omega)$, where $\rho(\Omega)$ is the mass-density distribution and $H(\Omega)$ stands for the height ($H(\Omega) > 0$) and depth ($H(\Omega) < 0$) of the mass distribution measured with respect to the mean sphere of radius R_L. For the topographic masses R_L is equal to R.

Interestingly, the spherical harmonic approach is insensitive to the vertical extension $H(\Omega)$, as it accounts only for the magnitude of the mass given by the surface density functions $\kappa_p(\Omega)$, which are expressed by the product of $\rho(\Omega)$ with $H(\Omega)^p/R^{p-1}$ ($p \in N^+$). This means that $H(\Omega)$ in Eq. (7) can be replaced by any (arbitrary) height function as long as $\rho(\Omega)$ is changed respectively in order to obtain the same value of the product above. Therefore, $H(\Omega)$ can simply be replaced by an equivalent rock height $H^{er}(\Omega)$ together with the constant reference density ρ_0 (see below). The advantage of using $H^{er}(\Omega)$ within the spherical harmonic approach is the possibility to combine mass distributions with different density functions to one single mass distribution with constant density.

In order to account for the vertical extension using the spherical harmonic approach, the magnitude of $H(\Omega)$ has to be limited, thus dividing a given mass distribution into several layered structures. The vertical dimension of the mass distribution is then accounted for by different upward continuation factors $(R_L/R)^{n+2}$ in Eq. (5). This property of the spherical harmonic approach is the reason for (theoretical and practical) differences between this approach and the numerical integration approach, which properly accounts for the 3-D structure of a given mass distribution.

The spherical harmonic coefficients of the gravitational potential [Eq. (5)] can be interpreted as an infinite sum of spherical harmonic coefficients representing surface density functions of the type given in Eq. (7). However, the spherical harmonic approach does not give the same results as obtained by condensing the regarded masses on the mean sphere with radius R_L (like in Helmert's condensation methods). In spherical approximation, the surface density of the condensed masses is given by (e.g. Kuhn 2002)

$$\kappa^c(\Omega) = \rho(\Omega)H(\Omega)\left(\frac{R}{R_L}\right)^2\left[1+\frac{H(\Omega)}{R_L}+\frac{H(\Omega)^2}{3R_L^2}\right] \tag{8}$$

Using the spherical harmonic expansion of $\kappa^c(\Omega)$ Eq. (5) can be expressed (here up to third order) by

$$\bar{V}_{nm} = \mu_n\left[\bar{\kappa}^c_{nm} + \frac{n}{2}\bar{\kappa}_{2,nm} + \frac{n-1}{6}\bar{\kappa}_{3,nm}\right]. \tag{9}$$

Equation (9) obviously is different to Eq. (5) showing that the spherical harmonic approach results in gravitational potential with higher spectral power than the gravitational potential caused the condensed masses using the surface density in Eq. (8).

3 Data

The numerical results presented in this paper are based on the global 5-arc-min x 5-arc-min DEM JGP95E (Lemoine et al. 1998), which classifies the terrain by six different types (1: dry land below MSL (mean sea level), 2: lake, 3: oceanic ice shelf, 4: ocean, 5: glacier ice, 6: dry land above MSL). These different mass distributions have been converted into equivalent rock heights with respect to the constant reference density of $\rho_0 = 2670$ kg/m^3.

Equivalent rock heights are based on the concept of mass equality between a given mass distribution described here by the lateral variable height and density functions $H(\Omega)$ and $\rho(\Omega)$, respectively, and the mass distribution described by the equivalent rock height $H^{er}(\Omega)$ and constant reference density ρ_0. In spherical approximation, the mass equality is given by

$$\frac{\rho(\Omega)}{R^2}\int_{r=R}^{R+H(\Omega)} r^2 dr = \frac{\rho_0}{R^2}\int_{r=R}^{R+H^{er}(\Omega)} r^2 dr \tag{10}$$

From Eq. (10), the equivalent rock height is given by

$$H^{er}(\Omega) = \sqrt[3]{\frac{\rho(\Omega)}{\rho_0}\left[(R+H(\Omega))^3 - R^3\right] + R^3} - R\,. \tag{11}$$

In Fig. 2 the global distribution of the equivalent rock height derived from JGP95E is illustrated.

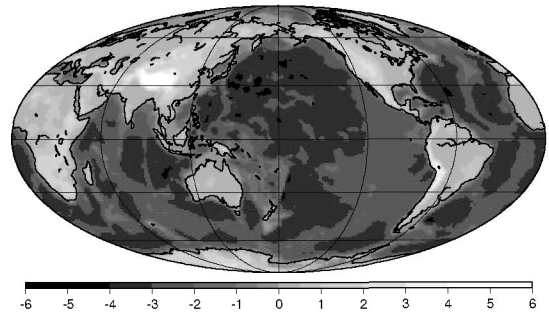


Fig. 2: Equivalent rock height derived from JGP95E [km].

The major differences between the original heights given by JGP95E and the equivalent rock heights occur in polar regions with their high ice masses.

It should be mentioned that the conversion to equivalent rock heights has been performed in order to deal with only one global mass distribution of constant density, which is more realistic to the actual mass distribution. It should not affect the numerical results as long as the same (equivalent rock) heights are used within both approaches.

4 Numerical Results

In order to get insight into the practical differences between the *discretised* numerical integration and spherical harmonic approaches, the effect on gravitational potential caused by the global topographic masses will be determined for both. Here, the topographic masses are defined by the positive equivalent rock heights (section 3) together with the constant mass-density $\rho_0 = 2670$ kg m^{-3}.

In the following, all computations are performed in a spherical approximation, thus the topographic masses are assumed to be located directly on the mean sphere of radius R. The spherical approximation has been chosen here in order to avoid any ellipsoidal effects in the numerical results.

Newton's integral has been evaluated using the *discretised* numerical integration and spherical harmonic approach given by Eqs. (2) and (3), respectively. Both approaches have been used in this paper up to third order of the corresponding Eqs. (2) and (5). A maximum degree of N=1440 has been used in the spherical harmonic approach. The 5-arc-min resolution of the equivalent rock heights (section 3) have been averaged to 15-arc-min in order to be consistent with the spatial resolution (at the equator) corresponding to the maximal degree N=1440.

The evaluation has been done globally on a 1-degree x 1-degree grid. The horizontal positions of the 1-degree x 1-degree grid elements are exactly collocated with the grid elements of JGP95E. Three different heights of the computation points have been used:

(1) Evaluation points are located on the Earth's surface defined here by the equivalent rock heights of JGP95E (section 3) over continents and H = 0 over oceans (on ocean surface).
(2) Evaluation points are located 4 km above the Earth's surface defined in item (1). This situation corresponds to a typical flight level in airborne gravimetry.
(3) Evaluation points are located on a sphere 400 km above the mean Earth's radius R corresponding to the height of a low-orbit satellite.

The gravitational potential derived by *discretised* numerical integration using spherical tesseroids is illustrated in Figs. 3 to 5 for each of the cases (1) to (3) mentioned above.

The spatial patterns in Figs. 3 to 5 are very similar to each other and roughly correlated with topography. The minimum and maximum are located in the central Pacific and Himalayas, respectively. A summary of all min./max. values of the cases (1) to (3) for the discretised numerical integration and spherical harmonic approach is given in Tab. 1.

For computation points located 400 km above the mean Earth radius [case (3)] the effect on gravitational potential is smaller and smoother than for the cases (1) and (2). This is due to the decrease (with $1/l$) of the gravitational potential effect with the distance to the source masses. Furthermore, this property explains the smoother pattern in Fig. (5) as Newton's integral acts like a low-pass filter for distant computation points, which can also be seen by the upward continuation term $(R_L/R)^{n+2}$ in Eq. (3). The spherical harmonic approach gives the same spatial pattern for the cases (1) to (3) as illustrated in Fig. 3 to 5 (not illustrated) but with slightly different minimum and maximum values (Table 1).

Tab. 1: Min./max. values for gravitational potential derived by discretised numerical integration and spherical harmonic approach.

	Num. integration		Spherical harmonic	
[m^2/s^2]	Min.	Max.	Min.	Max.
case (1)	1812.0	8101.5	1828.5	8114.0
case (2)	1811.4	8079.8	1827.9	8089.2
case (3)	1756.1	6221.0	1772.1	6232.5

Tab. 2: Min./max. values of differences between the numerical integration and spherical harmonic approach.

[m^2/s^2]	Min.	Max.
case (1)	+2.796	+33.334
case (2)	+0.082	+27.874
case (3)	+10.123	+26.759

More interesting are the differences between the numerical results of the *discretised* numerical integration and spherical harmonic approach (Table 2). For each of the three cases described above, the differences are illustrated in Figs. 6 to 8.

The differences are always positive and reach maximum values of more then 33 m^2/s^2. This means

smaller magnitudes are obtained by the *discretised* numerical integration than the spherical harmonic approach. While the differences of cases (1) and (2) show a negative correlation with topography (Figs. 6 and 7), this behaviour vanishes in case (3) (Fig. 8). Especially in the latter case, the differences show a rather systematic behaviour. As for the gravitational potential, the differences for case (3) also show a smoother spatial pattern with smaller magnitudes than those of cases (1) and (2) (see also Table 2).

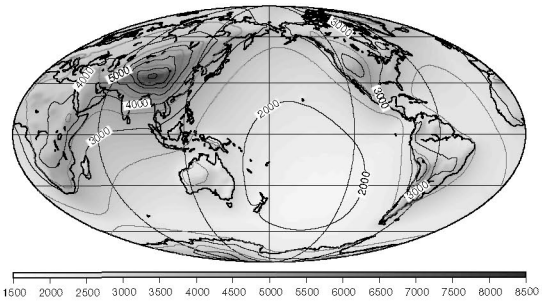


Fig. 3: Effect on gravitational potential due to global top. masses. Case 1: Evaluation points are on the Earth's surface. Contour interval [500 m^2/s^2].

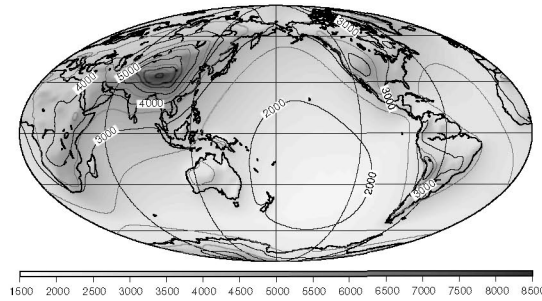


Fig. 4: Effect on gravitational potential due to global top. masses. Case 2: Evaluation points 4 km above the Earth's surface. Contour interval [500 m^2/s^2].

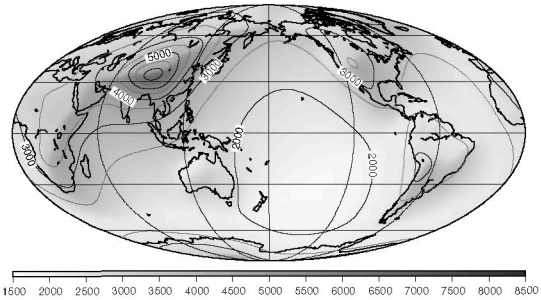


Fig. 5: Effect on gravitational potential due to global top. masses. Case 3: Evaluation points are at H = 400 km. Contour interval [500 m^2/s^2].

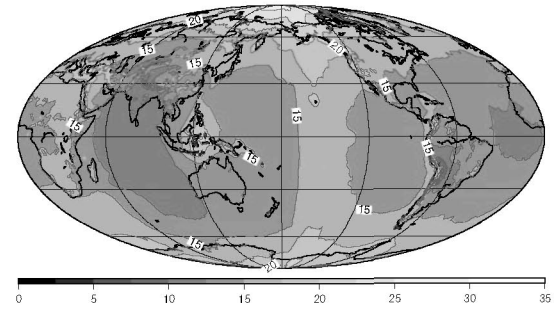


Fig. 6: Differences between the discretised num. integration and spherical harmonic approach. Case 1: Evaluation points are on the Earth's surface. Contour interval [2.5 m^2/s^2].

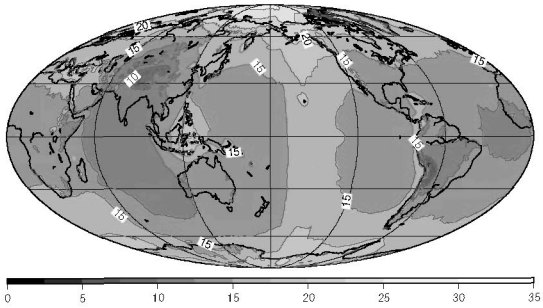


Fig. 7: Differences between the discretised num. integration and spherical harmonic approach. Case 2: Evaluation points 4 km above the Earth's surface. Contour interval [2.5 m^2/s^2].

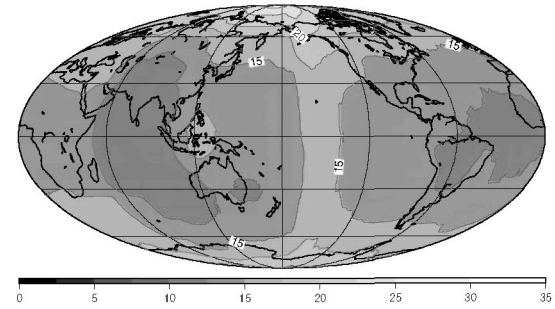


Fig. 8: Differences between the discretised num. integration and spherical harmonic approach. Case 3: Evaluation points are at H = 400 km. Contour interval [2.5 m^2/s^2].

5 Conclusions

It has been shown that the evaluation of Newton's integral by a *discretised* numerical integration and spherical harmonic approach leads to significant differences. In general the results for the spherical harmonic approach show higher magnitudes than the results of the *discretised* numerical integration. The differences reach maximum values of more than 33 m^2/s^2 at the Earth's surface for the gravitational potential caused by the global topographic masses. This corresponds roughly to a vertical change in equipotential surfaces of about 3 m (cf. Bruns formula). However, the effect should be about a magnitude smaller if the effect of the topographic masses is compensated by additional mass distributions below the geoid (e.g. isostatic compensation masses), which is still in the maximum range of several decimetres.

Presumably the differences are caused by the spherical harmonic approach, which does not account properly for the vertical extension of the given mass distribution. This affects to a greater extent evaluation points located on or near (e.g. several kilometres above) the Earth's surface. For longer distances the effect diminishes but remains still significant for points 400 km above the Earth's surface. Equally important is the spatial structure of the differences, which shows a rather systematic behaviour (all differences are positive). Therefore the spherical harmonic approach should not be used for precise global/local gravity field modelling.

For global gravity field studies (e.g. Kuhn and Featherstone 2003b, this issue) the spherical harmonic approach is well suited due to the global base functions used. Furthermore for low to medium degrees (say $N < 360$) it generally provides a saving in computation time with respect to the numerical integration. However, for very high degrees ($N > \sim 2700$) the spherical harmonic approach suffers from numerical instabilities (Holmes and Featherstone 2002b). In this case and for local studies, the (*discretised*) numerical integration approach has to be used.

Acknowledgements This study was partly funded by the Australian Research Council through the grant A00001127 and a small discovery grant from Curtin University of Technology.

References

Anderson EG (1976): The effect of topography on solutions of Stokes's problem. UNISURV Report S14, University of New South Wales, Kensington, Australia, pp. 252.

Heiskanen WA, Moritz H (1967) Physical Geodesy, Freeman, San Francisco, 364 pp.

Hildebrand FB (1974) Introduction to numerical analysis. McGraw-Hill, New York (reprinted 1987 by Dover Publications, New York, 669 pp).

Holmes SA, Featherstone WE (2002) A unified approach to the Clenshaw summation and the recursive computation of very high degree and order normalised associated Legendre functions, J Geod **76**: 279-299.

Kuhn M (2002) On the indirect effect of Helmert's condensation methods. International Geoid Service Bulletin **13**: 50-61.

Kuhn M, Featherstone WE (2003a) On the optimal spatial resolution of crustal mass distributions for forward gravity field modelling. In Tziavos IN (ed) Gravity and Geoid 2002, Proceedings of the 3rd meeting of the International Gravity and Geoid Commission, Thessaloniki, Greece, 195-200.

Kuhn M, Featherstone WE (2003b, this issue) Construction of a synthetic Earth gravity model by forward gravity modelling. Paper presented at the General Assembly of the International Union of Geodesy and Geophysics, June-July, 2003, Sapporo, Japan.

Lambeck K (1988) Geophysical Geodesy – The slow deformation of the Earth. Univ. Press, Oxford, 718 pp.

Lemoine FG and 14 other authors (1998) The development of the NASA GSFC and National Imaginary and Mapping Agency (NIMA) geopotential model EGM96. Rep. NASA/TP-1998-206861, National Aeronautics and Space Administration, Maryland, 575 pp.

Mader K (1951) Das Newtonsche Raumpotential prismatischer Körper und seine Ableitung bis zur dritten Ordnung. Österreichische Zeitschrift für Vermessungswesen, Sonderheft 11.

Nagy D (1966) The gravitational attraction of a right rectangular prism. Geophysics **31**: 362-371.

Nagy D, Papp G, Benedek J (2000) The gravitational potential and its derivatives for the prism, J Geod **74**: 552-560.

Nagy D, Papp G, Benedek J (2002) Corrections to "The gravitational potential and its derivatives for the prism", J Geod **76**: 475.

Ramillen G (2002) Gravity/magnetic potential of uneven shell topography, J Geod **76**: 139-149.

Seitz K, Heck B (2003) Efficient calculation of topographic reductions by the use of tesseroids. Presentation at the joint assembly of the EGS, AGU and EUG, Nice, France, 6-11 April 2003.

Wavelet evaluation of some singular geodetic integrals

QUANWEI LIU[1] AND MICHAEL G. SIDERIS

Department of Geomatics Engineering, University of Calgary, 2500 University Drive, Calgary, Alberta T2N 1N4, Canada

e-mail: sideris@ucalgary.ca; Tel: 1-403-220-4985; Fax: 1-403-284-1980
[1]Current address: Applanix Corporation, 85 Leek Crescent, Richmood Hill, ON, L4B 2B3, CANADA

Abstract. The paper presents a method of evaluating some singular geodetic integrals using wavelets. The wavelet transform is a powerful tool in evaluating some singular geodetic integrals. Due to its localization properties in both of the time (space) and frequency (scale) domains, and because the kernels of some geodetic integrals have singular points and decay smoothly and quickly away from the singularities, many numbers of wavelet transform coefficients of the kernels become zeros or negligible, and only small number of wavelet transform coefficients are significant. It is thus possible to significantly compress the kernels of these integrals in a wavelet basis by neglecting the zero coefficients and the small coefficients below a certain threshold. Therefore, wavelets provide a convenient way for efficiently evaluating these integrals in terms of fast computation and savings of computer memory. In this contribution, a modified algorithm for the wavelet evaluation of Stokes integrals is presented. And the same modified algorithm is applied to the evaluation of Vening Meinesz and terrain corrections integral, whose kernels have stronger singularities than of Stokes' kernel. Numerical examples illustrate the efficiency and accuracy of the wavelet methods.

Key words: *Wavelets – Kernel Compression – Stokes - Vening Meinesz- Terrain Corrections – Singularity – Computation Efficiency*

1 Introduction

During the last decade, the wavelet transform has proven to be a valuable tool in many application fields. Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functional. Due to wavelets are of local nature in both frequency (scale) and time (space), they can thus easily "detect" local features of the represented functions. This localization feature makes many functional and operators "sparse" when transformed into the wavelet domain. With multiresolution analysis (Mallat 1997), we can retain the large and important coefficients, and remove the small and unimportant coefficients whose absolute values are below a certain threshold.

The localization property of wavelets can be utilized more efficiently if the functional, which is represented in a wavelet basis, contains singular points or lines, since the functional can then be compressed with a larger compression ratio without loss of essential information. It is the sparse representation of geodetic singular integral operators in a wavelet basis that makes the application of wavelets in geodesy attractive (Benciolini, 1994; Barthelmes et al. 1994). Salamonowicz (1999) applied the two-dimensional wavelet transform to the fast computation of the singular Stokes integral point by point by using a special algorithm developed in Beylkin et al. (1991). In this paper we modify the algorithm and applied it to compute the Stokes and Verning Meinesz's and terrain correction integrals one meridian by one meridian in local or regional basis by the three-dimensional wavelet transform. We also include a comparison between the wavelet method with different compression ratios and the 2D planar FFT methods in term of computational efficiency and accuracy.
Due to the limit of space, the reader is referred to Liu and Sideris (2003) for an overview of the wavelet theory.

3 The wavelet method

The reason to employ wavelets to evaluate geodetic integrals is that the kernels of some geodetic integrals have singular points and they decay smoothly and quickly away from the singularities in the non-singular parts of the integral. Because a large number of wavelets coefficients become zeros, there is only a small number of coefficients which are significant due to the localized nature of wavelets. And therefore the integral kernel can be sparse and compressed in a wavelet basis by neglecting these zero coefficients and the small coefficients below a certain threshold. Due to the sparse representation of the kernel in a wavelet basis and because of the fast wavelet transform algorithm the geodetic integrals can be computed efficiently by using wavelets.

3.1 The three integrals: Stokes integral, Vening Meinesz integral and terrain corrections

If the gravity anomaly data $\Delta g[n,m]$ are given in a limited area E, we take the integration area as a plane and define a local rectangular coordinate system, the geoidal undulation N can be computed by the Stokes formula in planar approximation (Schwarz et al. 1990)

$$N(x_2,y_2)=\frac{1}{2\pi\gamma}\iint_E \Delta g(x_1,y_1)\frac{1}{[(x_2-x_1)^2+(y_2-y_1)^2]^{1/2}}dx_1dy_1 \quad (1)$$

(x_1,y_1) and (x_2,y_2) are local Cartesian coordinates of the running point and computation point, respectively and, γ is the normal gravity.
From formula (1) we can immediately recognize that the kernel

$$K(x_1,y_1,x_2,y_2)=\frac{1}{[(x_2-x_1)^2+(y_2-y_1)^2]^{1/2}}, \quad (2)$$

has a singularity at each computation point, and it decreases smoothly away from the singular point. This is the very reason that wavelets can be applied to compress the Stokes kernel and efficiently compute the integral.

Another commonly used singular geodetic integral is the Vening Meinesz integral to determine deflections of the vertical. The integral in planar approximation may be written as (3) (Schwarz et al. 1990)

$$\begin{Bmatrix}\xi(x_2,y_2)\\ \eta(x_2,y_2)\end{Bmatrix}=\frac{1}{2\pi\gamma}\iint_E \Delta g(x_1,y_1)\frac{1}{[(x_2-x_1)^2+(y_2-y_1)^2]^{3/2}}\begin{Bmatrix}y_2-y_1\\ x_2-x_1\end{Bmatrix}dx_1dy_1 \quad (3)$$

The integral formula giving the terrain correction N at a point P on the plane reference surface may be written as (Sideris 1985)

$$N(x_2,y_2)=\frac{1}{2}G\rho\iint_E \frac{[h(x_1,y_1)-h(x_2,y_2)]^2}{[(x_2-x_1)^2+(y_2-y_1)^2]^{3/2}}dx_1dy_1 \quad (6)$$

(x_1,y_1) and (x_2,y_2) are local Cartesian coordinates of the running point and computation point, respectively and, G is the gravitional constant, ρ is the mean density of the topographical masses (assumed constant) and h is the height of the points.

From formula (3) we can also recognize that the kernel at each meridian

$$\begin{Bmatrix}K_\xi\\ K_\eta\end{Bmatrix}=\frac{1}{[(x_c-x_1)^2+(y_2-y_1)^2]^{3/2}}\begin{Bmatrix}y_2-y_1\\ x_c-x_1\end{Bmatrix} \quad (4)$$

has a singular line at the computation meridian, and it decreases away from the singular line.

In order to further examine the singularity of the above kernel, we rewrite it as

$$\begin{Bmatrix}K_\xi\\ K_\eta\end{Bmatrix}=\frac{1}{[(x_c-x_1)^2+(y_2-y_1)^2]^{1/2}}\frac{\begin{Bmatrix}y_2-y_1\\ x_c-x_1\end{Bmatrix}}{[(x_c-x_1)^2+(y_2-y_1)^2]} \quad (5)$$

At the computation point, the denominator $[(x_c-x_1)^2+(y_2-y_1)^2]$ approaches zeros much faster than the numerator (x_c-x_1), or (y_2-y_1) of the second part of the kernel. Therefore the kernel has stronger singularity than the kernel of the Stokes integral, and it decreases much more rapidly with increasing distance $s=[(x_c-x_1)^2+(y_2-y_1)^2]^{1/2}$. This means that the kernels K_ξ and K_η can be compressed significantly more than the Stokes kernel K since the larger wavelet coefficients are more concentrated around the singularity line due to the strong singularity and the rapid decreasing of the kernel.

From formula (6) we can immediately recognize that the kernel

$$K_T(x_1,y_1,x_2,y_2)=\frac{1}{[(x_2-x_1)^2+(y_2-y_1)^2]^{3/2}}, \quad (7)$$

has a singularity at each computation point, and it decreases much faster than that of the Stokes's. Therefore the kernel has stronger singularity than the kernel of the Stokes integral, and it decreases much more rapidly. This means that the kernels of terrain corrections can be compressed significantly more than the Stokes kernel since the larger wavelet coefficients are more concentrated around the singularity line due to the strong singularity and the rapid decreasing of the kernel.

Note that in practical computation, K_T is zero in the computation point since a line mass at the computation point itself contributes nothing to the terrain correction (Sideris 1985).

Since our paper "Wavelet evaluation of the Stokes and Vening Meinesz integrals" was accepted for publication in Journal of Geodesy (Liu and Sideris 2003). The following section will focus on wavelet evaluation of terrain corrections.

3.2 Wavelet evaluation

Formula (7) can be expressed as

$$C(x_2,y_2)=\frac{1}{2}G\rho\iint_E K(x_1,y_1,x_2,y_2)H(x_1,y_1)^2dx_1dy_1 - G\rho H(x_2,y_2)\iint_E K(x_1,y_1,x_2,y_2)H(x_1,y_1)dx_1dy_1+\frac{1}{2}G\rho H(x_2,y_2)^2\iint_E K(x_1,y_1,x_2,y_2)dx_1d \tag{8}$$

Due the similarity of the three integrals in the above terrain formula, in the following only the algorithm for computing the second integral is developed for the sake of simplicity. The fast algorithm for computing the two-dimensional wavelet transform coefficients of the height at scale 2^j in their continuous forms is:

$$b^1_{pq(j)} = \iint \Delta g(x_1,y_1)\psi_{p(j)}(x_1)\psi_{q(j)}(y_1)dx_1dy_1 \tag{9.1}$$

$$b^2_{pq(j)} = \iint \Delta g(x_1,y_1)\psi_{p(j)}(x_1)\phi_{q(j)}(y_1)dx_1dy_1 \tag{9.2}$$

$$b^3_{pq(j)} = \iint \Delta g(x_1,y_1)\phi_{p(j)}(x_1)\psi_{q(j)}(y_1)dx_1dy_1 \tag{9.3}$$

$$b^4_{pq(j)} = \iint \Delta g(x_1,y_1)\phi_{p(j)}(x_1)\phi_{q(j)}(y_1)dx_1dy_1 \tag{9.4}$$

Where $\psi_{p(j)}$, $\psi_{q(j)}$ and $\phi_{p(j)}$, $\phi_{q(j)}$ represent the orthogonal wavelet and scaling functions at the dyadic intervals at the scale 2^j when referring to a running point in x and y directions, respectively. Δg represents the height and is renamed as b^4 for the sake of easy notation.

In the following we apply the wavelet transform and reconstruction to the kernel. If the terrsin corrections at all of the computing points are computed at once, then we need to apply the four-dimensional wavelet transform to the kernel. However, as Salamonowicz (1999) pointed out, the four-dimensional wavelet transform of the kernel can hardly be compressed from the point of view of computer programming, so he applied the two-dimensional wavelet transform to the kernel so that the transformed kernel is compressed in a two-dimensional basis, and the geoidal undulation is computed point by point. We believe that the pointwise computation of the geoidal undulation takes too much time, and it undermines the efficiency of wavelet application in computing the integral. Here we attempt to make a compromise; we want to compute the geoidal undulation meridian by meridian instead of computing all the points at once or computing point by point. For this reason, we need to apply the three-dimensional wavelet transform to the kernel, and the kernel is compressed in a three-dimensional basis.

The kernel is the same as formula (8), but the coordinate component x_2 of the computation point becomes constant at each meridian. Subsequently, we denote the kernel at each meridian as

$$K(x_1,y_1,y_2)=\frac{1}{[(x_c-x_1)^2+(y_2-y_1)^2]^{3/2}} \tag{10}$$

where the constant x_c is the x-coordinate of every computation point in the direction of the computation meridian. The kernel has a singular line at the computation meridian.

Analogously to the continuous forms of the two-dimensional wavelet transform of the height as shown in equation (9), the continuous forms of the three-dimensional wavelet transform of the kernel at each meridian at the scale 2^j can be expressed as

$$d^1_{pqs(j)} = \iiint K(x_1,y_1,y_2)\psi_p(x_1)\psi_q(y_1)\psi_s(y_2)dx_1dy_1dy_2$$

$$d^2_{pqs(j)} = \iiint K(x_1,y_1,y_2)\psi_p(x_1)\psi_q(y_1)\phi_s(y_2)dx_1dy_1dy_2$$

$$d^3_{pqs(j)} = \iiint K(x_1,y_1,y_2)\psi_p(x_1)\phi_q(y_1)\psi_s(y_2)dx_1dy_1dy_2$$

$$d^4_{pqs(j)} = \iiint K(x_1,y_1,y_2)\phi_p(x_1)\psi_q(y_1)\psi_s(y_2)dx_1dy_1dy_2$$

$$d^5_{pqs(j)} = \iiint K(x_1,y_1,y_2)\phi_p(x_1)\phi_q(y_1)\psi_s(y_2)dx_1dy_1dy_2$$

$$d^6_{pqs(j)} = \iiint K(x_1,y_1,y_2)\phi_p(x_1)\psi_q(y_1)\phi_s(y_2)dx_1dy_1dy_2$$

$$d^7_{pqs(j)} = \iiint K(x_1, y_1, y_2)\psi_p(x_1)\phi_q(y_1)\phi_s(y_2)dx_1dy_1dy_2$$

$$d^8_{pqs(j)} = \iiint K(x_1, y_1, y_2)\phi_p(x_1)\phi_q(y_1)\phi_s(y_2)dx_1dy_1dy_2 \tag{11}$$

where ψ_s, and ϕ_s represent the orthogonal wavelet and scaling functions at the dyadic interval at the scale 2^j referring to a computation point in the meridian directions, i.e., the y directions. The above three-dimensional wavelet coefficients can also be computed recursively with the fast algorithm similar to the two-dimensional fast algorithm shown in the last section.

After computing all of these 7J+1 coefficients, where J is the largest scale ($J = \log_2^n$, n is the total number of data points), the original kernel at each meridian is recovered from this wavelet representation by the following reconstruction formula (12)

$$\begin{aligned} K(x_1, y_1, y_2) = & \sum_j\sum_p\sum_q\sum_s d^1_{pqs(j)}\psi_p(x_1)\psi_q(y_1)\psi_s(y_2) + \sum_j\sum_p\sum_q\sum_s d^2_{pqs(j)}\psi_p(x_1)\psi_q(y_1)\phi_s(y_2) + \\ & \sum_j\sum_p\sum_q\sum_s d^3_{pqs(j)}\psi_p(x_1)\phi_q(y_1)\psi_s(y_2) + \sum_j\sum_p\sum_q\sum_s d^4_{pqs(j)}\phi_p(x_1)\psi_q(y_1)\psi_s(y_2) \\ & \sum_j\sum_p\sum_q\sum_s d^5_{pqs(j)}\phi_p(x_1)\phi_q(y_1)\psi_s(y_2) + \sum_j\sum_p\sum_q\sum_s d^6_{pqs(j)}\phi_p(x_1)\psi_q(y_1)\phi_s(y_2) + \\ & \sum_j\sum_p\sum_q\sum_s d^7_{pqs(j)}\psi_p(x_1)\phi_q(y_1)\phi_s(y_2) + \sum_p\sum_q\sum_s d^8_{pqs(J)}\phi_p(x_1)\phi_q(y_1)\phi_s(y_2) \end{aligned} \tag{12}$$

$$\begin{aligned} N(x_c)_2 = & \frac{G\rho h(x_2 - y_2)}{2}\iint[\sum_j\sum_p\sum_q\sum_s d^1_{pqs(j)}\psi_p(x_1)\psi_q(y_1)\psi_s(y_2) + \sum_j\sum_p\sum_q\sum_s d^2_{pqs(j)}\psi_p(x_1)\psi_q(y_1)\phi_s(y_2) + \\ & \sum_j\sum_p\sum_q\sum_s d^3_{pqs(j)}\psi_p(x_1)\phi_q(y_1)\psi_s(y_2) + \sum_j\sum_p\sum_q\sum_s d^4_{pqs(j)}\phi_p(x_1)\psi_q(y_1)\psi_s(y_2) + \\ & \sum_j\sum_p\sum_q\sum_s d^5_{pqs(j)}\phi_p(x_1)\phi_q(y_1)\psi_s(y_2) + \sum_j\sum_p\sum_q\sum_s d^6_{pqs(j)}\phi_p(x_1)\psi_q(y_1)\phi_s(y_2) + \\ & \sum_j\sum_p\sum_q\sum_s d^7_{pqs(j)}\psi_p(x_1)\phi_q(y_1)\phi_s(y_2) + \sum_p\sum_q\sum_s d^8_{pqs(J)}\phi_p(x_1)\phi_q(y_1)\phi_s(y_2)]\Delta g(x_1, y_1)dx_1dy_1 \end{aligned} \tag{13}$$

$$\begin{aligned} N(x_c)_2 = & \frac{G\rho h(x_2, y_2)}{2}\times\sum\sum\sum(d^1_{pqs(1)}b^1_{pq(1)} + d^3_{pqs(1)}b^2_{pq(1)} + d^4_{pqs(1)}b^3_{pq(1)} + d^5_{pqs(1)}b^4_{pq(1)})\psi_s(y_2) + \\ & \frac{G\rho}{2}\times\sum\sum\sum(d^2_{pqs(1)}b^1_{pq(1)} + d^6_{pqs(1)}b^3_{pq(1)} + d^7_{pqs(1)}b^2_{pq(1)})\phi_s(y_2) + d^8_{pqs(J)}b^4_{pq(J)}\phi_s(y_2) \end{aligned} \tag{14}$$

The last term in formula(12) is the remaining coarse kernel approximation at scale 2^J. Substituting the reconstructed kernel into the second term of formula (8) yields the second integral of the terrain corrections in the computation meridian as shown in formula (13)

Interchanging the order of integration and summation, moving Δg into each integration and comparing formula (13) and formula (9), we obtain formula (14)

Anagously the first term $N(x_c)_1$ and third term $N(x_c)_3$ of formula (9) can also be computed as in the above procedure. And the final terrain correction is the sum of the three integrals.

Anagously the Stokes' and Vening Meinesz integrals can be computed using wavelet in the same way as computing the terrsin corrections as discussed in the above. For the sake of saving the space, we are not going to the details. The interested reader is referred to (Liu and Sideris 2003).

4 Numerical Examples

Considering the reason that the wavelets are introduced in physical geodesy is their expected high efficiency of

evaluation in terms of savings of memory and computation speed, it is obviously necessary to compare wavelets with the highly developed and widely accepted efficient FFT approach in evaluating the three integrals. The three integrals were computed in the Rocky Mountains of British Columbia, Canada. The data used is a grid of 64 by 64 of 5' point gravity anomalies in the area bounded by latitude 49°N to 54.4°N and longitude 236°E to 241.4°E. The same grid of heights in the same area is also available. Table 1 gives the statistical information of the free-air gravity anomalies and heights.

Table 1. Statistics of gravity anomalies in mGal and heights in meter

data	max	min	mean	rms	σ
height	2823	0	1223	1232	147
gravity	132	-159	7	49	49

Both the geoidal undulations and deflections were computed by pointwise integration on the plane, which serves as the reference, and by 2D FFT and wavelets on the plane. the terrain corrections were computed using linemass method by pointwise integration on the plane, which serves as the reference, and by 2D FFT and wavelets on the plane. The kernels were pre-computed first before applying pointwise integration and wavelets. Daubechies wavelet with vanishing moment 4 is chosen for transforming the kernel because it maximizes the number of zero coefficients and minimizes the number of high amplitude coefficients. Different threshold values for the wavelet coefficients of the kernel, which result in different kernel compression ratios, were employed. And the coefficients, whose absolute values were larger than the threshold, and their indexes were saved into two separate vectors in the first stage. They were retrieved to compute the three integrals in the second stage. The wavelet methods with different compression rations are compared with pointwise integration and FFT in terms of computational efficiency and accuracy.

Table 2, Table 3 and Table 4 shows the statistical information of the geoidal undulations, the meridian component of the deflections and terrain corrections computed by pointwise integration on the plane, by 2D FFT on the plane, and by wavelets with different kernel compression ratios. The values following the names of wavelet methods in the first column are the wavelet coefficient thresholds for compressing the kernels. From the three tables, one may wonder that as the data elements for wavelet methods W30, W26, W22, W9 and W7 are smaller than that for FFTs, but why they needs more computations than FFTs do. This can be answered by the fact the FFTs compute all the points at once; while wavelets compute one meridian by one meridian.

Table 2. Comparison of the geoidal undulations from different methods

Evaluation method	max	min	mean meter	rms	σ	CPU sec	gravity elements	kernel elements
Pointwise (PT)	6.945	-2.227	2.118	2.774	1.792	12.3	64×64	64×64
2D FFT (FFT)	6.945	-2.227	2.118	2.774	1.792	2.4	128×128	128×128
Wavelet (W30) 0.05	6.942	-2.228	2.118	2.774	1.792	5.6	64×64	2×1146
Wavelet (W26) 0.3	6.908	-2.233	2.112	2.774	1.799	5.0	64×64	2×1024
Wavelet (W22) 0.4	6.863	-2.370	2.012	2.674	1.761	4.7	64×64	2×784
PT- FFT	0.000	0.000	0.000	0.000	0.000			
PT-W30	0.008	-0.007	0.000	0.002	0.002			
PT-W26	0.231	-0.138	0.006	0.043	0.043			
PT-W22	0.684	-0.295	0.106	0.215	0.187			

Table 3. Comparison of the ξ deflection from different methods

Evaluation method	max	min	mean arc-sec	rms	σ	CPU sec	gravity elements	kernel elements
Pointwise (PT)	21.4	-24.7	0.2	5.1	5.1	12.8	64×64	64×64
2D FFT (FFT)	21.4	-24.7	0.2	5.1	5.1	2.7	128×128	128×128
Wavelet (W9) 0.004	21.3	-24.7	0.1	5.1	5.1	3.0	64×64	2×368
Wavelet (W3) 0.04	19.0	-23.7	-0.0	4.8	4.8	2.6	64×64	2×122
Wavelet (W1) 0.4	14.7	-17.2	0.0	3.9	3.9	2.4	64×64	2×42
PT- FFT	0.0	-0.0	0.0	0.0	0.0			
PT-W9	0.9	-0.7	0.1	0.4	0.3			
PT-W3	3.8	-4.2	0.2	1.2	1.0			
PT-W1	7.6	-8.0	-0.1	1.9	1.9			

Table 4. Comparison of the terrain corrections from different methods

Evaluation method	max	min	mean mGal	rms	σ	CPU sec	height elements	kernel elements
Pointwise (PT)	15.8	0.1	1.8	2.5	1.8	14.3	64×64	64×64
2D FFT (FFT)	15.8	0.1	1.8	2.5	1.8	3.4	128×128	128×128
Wavelet (W7) 0.0001	15.6	0.0	1.7	2.4	1.8	4.1	64×64	2×286
Wavelet (W3) 0.001	14.6	-0.0	1.5	2.2	1.7	2.8	64×64	2×122
Wavelet (W0.7) 0.01	13.0	-0.3	1.3	2.0	1.5	2.2	64×64	2×28
Wavelet (W0.1) 0.05	12.6	-0.9	0.8	1.4	1.1	1.9	64×64	2×4
PT- FFT	0.0	0.0	0.0	0.0	0.0			
PT-W7	0.7	0.0	0.1	0.1	0.0			
PT-W3	1.5	-0.0	0.2	0.2	0.0			
PT-W0.7	3.0	-0.4	0.4	0.4	0.1			
PT-W0.1	8.2	-1.2	0.9	0.9	0.2			

5 Analysis and conclusions

From the theoretical developments and numerical experiments in this paper, it is seen that singular geodetic integrals of the Stokes and Vening Meinesz and terain corrections can be compressed in a wavelet basis. It is shown the Vening Meinesz integral and terrain correction integral can be evaluated more efficiently due to its stronger singularity. A geoid with mm, cm and dm accuracy can be computed by the wavelet method with a kernel compression ratio of 30%, 26% and 22%, respectively. Deflections of vertical with sub arc-second, one arc-second and two arc-seconds accuracy can be obtained by the wavelet method with a compression ratio of 9%, 3% and 1%. Terrain corrections with sub mGal, one mGal accuracy can be obtained by the wavelet method with a compression ratio of 3%, 0.7%. Compared with FFT techniques, it is shown that the wavelet evaluation of the Stokes and Vening Meinesz and terrain integrals needs much less memory allocation than the FFT methods. It is also shown that wavelets can be faster than FFT if one wants to obtain the deflections with one arc-second accuracy, but it is slower than FFT if one wants to compute the geoid undulations with cm accuracy. It is also shown that wavelets can be faster than FFT if one wants to obtain the terrain correction with one mGal accuracy. Therefore, when computing the deflections of the vertical and terrain corrections with an acceptable accuracy, the wavelet method is a wise alternative choice in place of FFT if one wants to save computer memory and computer time. However, the wavelet method is less efficient than the FFT method in terms of speed in evaluating geoidal undulations with an acceptable accuracy and therefore is not recommended in practical use if one does not want to scarify much accuracy for the savings of computer memory.

Acknowledgements

Support for this research was provided by research grants to Dr. Sideris from the Natural Sciences and Engineering Reserch Council of Canada (NSERC) and the Geomatics for Informed Decisions (GEOIDE) Network of Centres of Excellence.

References

Barthelmes F, Ballani L, Klees R (1994) On the Application of Wavelets in Geodesy. Geodetic Theory Today – Third Hotine-Marussi Symposium on Mathematical Geodesy – Symposium No. 114: 394-403. Springer Verlag, Berlin/New York

Benciolini B (1994) A Note on Some Possible Geodetic Applications of Wavelets. Section IV Bulletin IAG

Beylkin, G, Coifman R, Rokhlin V (1991) Fast Wavelet Transforms and Numerical Algorithms I. Communications in Pure and Applied Mat 44: 141-183

Daubechies I (1996) Where do wavelets come from? (A personal point of view). Proceedings of the IEEE, Special Issue on Wavelets 84 :510--513

Heiskanen WA, Moritz H (1967) Physical Geodesy. W. H. Freeman and Company, San Francisco

Liu QW, Li YC, Sideris MG (1997) Evaluation of deflections of the vertical on the sphere and the plane - a comparison of FFT techniques, Journal of Geodesy 71: 461-468

Liu QW and Sideris MG (2002) Wavelet evaluation of the Stokes and Verning Meinesz integrals, Journal of Geodesy 77: 345:356

Mallat S (1997) A Wavelet Tour of Signal Processing. Academic Press, San Diego

Salamonwicz PH (1999) A Wavelet Based Gravity Model with an Application to the Evaluation of Stokes' Integral. In: Sideris (ed.) International Association of Geodesy Symposium 123, Gravity, Geoid, and Geodynamics, pp85-90

Schwarz KP, Sideris MG, Forsberg R (1990) The Use of FFT Techniques in Physical Geodesy. Geophys. J. Int. 100: 485-514

Sideris MG (1985) A Fast Fourier Transform method for computing terrain corrections. Manu Geod 10: 66-73

Gravimetric Geoid Determination in the Andes

C. Tocho and G.Font
Facultad de Ciencias Astronómicas y Geofísicas, Paseo del Bosque s/n, 1900 La Plata, Argentina.

M.G. Sideris
Department of Geomatics Engineering, The University of Calgary, 2500 University Drive N.W., Calgary, Alberta, T2N 1N4, Canada.

Abstract. As a preliminary analysis for the development and evaluation of a precise gravimetric geoid for Argentina, different gravimetric geoid solutions were computed using different gravity reduction techniques in the roughest area of the country.

The rugged area bounded by latitude 20° S to 42° S and longitude 72° W to 67° W was chosen to compute the geoids using the following terrain reductions methods: Helmert's second condensation method, Rudzki's inversion method, the Airy-Heiskanen topographic-isostatic reduction, and the residual terrain model (RTM) method. The remove-restore method was employed using the most homogenised gravity database (with data blunders removed) referenced to a unified datum, the EGM96 global geopotential model and GTOPO30 digital elevation model, to take into account the short-wavelength component of the gravity signal. Stokes's formula was evaluated using the approximated spherical kernel by the two-dimensional fast Fourier transform algorithm.

The external accuracy of the gravimetric geoid models was evaluated by comparing them with undulations derived from GPS/levelling data. A four-parameter transformation was used to remove the systematic datum differences between the gravimetric geoid and the GPS/levelling undulations, and the possible long wavelength errors of the geoid

Keywords. Geoid, terrain reductions, Argentina

1 Introduction

The use of Stokes's formula in gravimetric geoid determination requires that the gravity anomalies represent boundary values on the geoid. This means that the measured gravity (usually taken on the surface of the Earth) must be reduced to the geoid and there must be no masses outside the geoid (Heiskanen and Moritz, 1967).

There are several gravity reductions used in physical geodesy: the Bouguer reduction, topo-isostatic reductions, the Rudzki inversion method, the second method of Helmert's condensation, etc. The Residual Terrain Model (Forsberg, 1984) is another type of reduction, which takes into account the high frequencies of the topography. RTM yields the quasigeoid, thus the correction between quasigeoid and geoid was applied to compare the results from the three methods with GPS/leveling data.

Theoretically, all reduction methods should lead to the same geoid if the gravity reductions were applied consistently (Heiskanen and Moritz, 1967) even though each reduction treats the topography in a different way.

2 Computational formulas for terrain effects

The four methods compared in this investigation use the remove-restore technique for the determination of the gravimetric geoid/quasigeoid. Each method handles the topography in a different way. The geoid or quasigeoid is calculated from Stokes's formula with gravity anomalies as input. Before applying this formula, gravity anomalies must be reduced, in the remove step

$$\Delta g = \Delta g_{FA} - \Delta g_T - \Delta g_{GM} \quad (1)$$

where Δg_{FA} is the free-air gravity anomaly, Δ_{GM} is the reference gravity anomaly computed from the geopotential model and Δg_T is the direct topographical effect on gravity, which depends on the reduction method used.

The gravimetric geoid is obtained, in the restore step, by

$$N = N_{GM} + N_{\Delta g} + N_{ind} \quad (2)$$

where N_{GM} is the reference geoidal undulation implied by the geopotential model, N_{ind} is the indirect effect on the geoid and depends on the reduction method used, and $N_{\Delta g}$ represents residual geoid computed with the residual gravity anomalies given in Eq. (1).

Different approximations to Stokes' kernel were investigated to compute residual geoid undulations in the spectral domain. Finally, it was decided to

employ the 2D-FFT spherical Stokes convolution to evaluate the kernel function (Strang Van Hess, 1990).

$$N_{\Delta g} = \frac{R\Delta\varphi\Delta\lambda}{4\pi\gamma} F^{-1}\{F(\Delta g \cos\varphi) F(\Delta\varphi, \Delta\lambda, \varphi_m))\} \quad (3)$$

where $N_{\Delta g}$ are the estimated residual gravimetric geoid heights or residual cogeoid heights.

The indirect effect on the geoid is

$$N_{ind} = \frac{\Delta T}{\gamma} \quad (4)$$

where ΔT is the change of the potential at the geoid due to the terrain reduction applied and γ is the normal gravity. The indirect effect on gravity, which reduces gravity anomalies from the geoid to the cogeoid should be added to Eq. 1 for Helmert's second method of condensation and Airy method, and can be expressed by (Heiskanen and Moritz, 1967) by

$$\delta\Delta g = 0.3086 N_{ind}[mGal] \quad (5)$$

The RTM method estimates the quasigeoid. The free-air gravity anomalies and Eq. 1 refer to the surface of the topography. Eq. 2 for height anomaly or quasigeoid determination can be formulated as

$$\zeta = \zeta_{GM} + \zeta_{\Delta g} + \zeta_{ind} \quad (6)$$

where ζ_{ind} represents the indirect effect on the quasigeoid for the RTM reduction.

In order to compare the results of this method with the other three reduction methods and to make comparisons with the GPS/leveling derived geoid, the quasigeoid in the RTM method is converted to geoid using the quasigeoid-geoid separation given in (Heiskanen and Moritz, 1967) as

$$N \approx \zeta + \frac{\Delta g_B}{\bar{\gamma}} H \quad (7)$$

where $\bar{\gamma}$ and Δg_B represent the mean normal gravity and Bouguer anomaly, respectively.

2.1 Helmert's second method of condensation

Helmert's second method of condensation condenses the topographic masses on a surface layer on the geoid and it can be view as a limiting case of Pratt-Hayford when the depth of condensation goes to zero. Before applying Stokes's formula, the indirect effect on gravity has been considered.

The terrain effect on gravity Δg_T is given by

$$\Delta g_T = -\delta\Delta g - c_P \quad (8)$$

where $\delta\Delta g$ is the indirect effect on gravity given by (Sideris and She, 1995) as

$$\delta\Delta_g = \frac{2\pi G\rho h^2}{\gamma} \quad (9)$$

and c_P is the classical terrain correction given by the following equation:

$$c_P = -G\iint_E \int_{h_P}^{h} \frac{\rho(x,y,z)(h_P - z)}{r^3(x_P - x, y_P - y, h_P - z)} dxdydz \quad (10)$$

where $\rho(x,y,z)$ is the topographical density at the running point and G is the gravitational constant.

The indirect effect on the geoid, up the second order is in planar approximation (Wichiencharoen, 1982)

$$N_{ind} = -\frac{\pi G\rho h_P^2}{\gamma} - \frac{G\rho}{6\gamma}\iint_E \frac{h^3 - h_P^3}{r_0^3} dxdy \quad (11)$$

where r_0 is the planar distance between computation point and data point.

2.2 Rudzki inversion method

The Rudzki inversion method is a reduction where the indirect effect is zero. Rudzki shifts all the topographic masses inside the geoid.

$$\Delta g_T = A_T - A_{inv} \quad (12)$$

where A_T is the attraction of the topographic masses, A_{inv} is the attraction of the inverted masses, with the density of the topographic masses being equal to the density of the inverted masses and the thickness of the inverted masses is equal to the height of the topography.

$$A_T = -G\iint_E \int_0^{h} \frac{\rho(x,y,z)(h_P - z)}{r^3(x_P - x, y_P - y, h_P - z)} dxdydz \quad (13)$$

$$A_{inv} = -G\iint_E \int_{h-h_P}^{-h_P} \frac{\rho(x,y,z)(h_P - z)}{r^3(x_P - x, y_P - y, h_P - z)} dxdydz \quad (14)$$

3.3 RTM method

The RTM reduction method takes into account the high frequencies of the topography. The effect of the topography above a long wavelength topographic surface is first removed and later restored.

The RTM gravity terrain effect, Δg_T is given by the approximate expression (Forsberg, 1984)

$$\Delta g_T = \Delta g_{RTM} \approx 2\pi G\rho(h - h_{ref}) - c_P \qquad (15)$$

where h is the topographic height given by a digital terrain model, h_{ref} is the height of a smooth mean reference surface and c_P is the classical terrain correction.

The RTM geoid effect is expressed in linear approximation as:

$$\zeta_T = \zeta_{RTM} = \frac{G\rho}{\gamma}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{h_{ref}}^{h}\frac{1}{r}dxdydz = \frac{G\rho(h-h_{ref})}{\gamma}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{r_0}dxdy \qquad (16)$$

where r_0 is the planar distance

2.4 Airy method

The object of an isostatic reduction of gravity, according to Airy-Heiskanen (AH) model is the regularization of the earth's crust (see Heiskanen and Moritz, 1967). The terrain effect Δg_T is

$$\Delta g_T = A_T - A_C \qquad (17)$$

$$A_T = -G\iiint_E^h \frac{\rho(x,y,z)(h_P - z)}{r^3(x_P - x, y_P - y, h_P - z)}dxdydz \qquad (18)$$

$$Ac = -G\iiint_E^{T-h_P} \frac{\Delta\rho(x,y,z)(h_P - z)}{r^3(x_P - x, y_P - y, h_P - z)}dxdydz \qquad (19)$$

where A_T is the attraction of the topographic masses above the geoid and A_c is the attraction of the compensated masses according AH model, t is the root height, h is the height of the topography, $\Delta\rho$ is the density contrast between the upper mantle and the crust, ρ is the crust density and T is the thickness of the Earth's crust.

The change of potential ΔT in equation (4) is for the isostatic reduction expressed by

$$\Delta T = T_T - T_C \qquad (20)$$

where T_T and T_C are the gravitational potentials of the topography and compensating masses respectively

$$T_T = -G\iiint_E^h \frac{\rho(x,y,z)}{r\,(x_P - x, y_P - y, h_P - z)}dxdydz \qquad (21)$$

$$T_C = -G\iiint_E^{T} \frac{\Delta\rho(x,y,z)}{r\,(x_P - x, y_P - y, h_P - z)}dxdydz \qquad (22)$$

3 Area under study and data availability

The area under study is located in the roughest area of Argentina, bounded by 20°S to 42°S in latitude and 67°W to 72°W in longitude.

The point gravity measurements, provided by different sources, were referenced to the International Standardization Net 1971 (I.G.S.N.71).

A total of 11132 measured gravity points, with a mean data spacing of approximately 12 km are used in the mountainous area The distribution of the gravity points is shown in Figure 1. Gravity anomalies are computed using the parameters of GRS80 according to the formulas given in Torge, 1989.

The reference gravity field is computed from the EGM96 geopotential model (Lemoine et al., 1998) complete to degree and order 360.

The global digital elevation model GTOPO30 with a horizontal grid spacing of 30 arc seconds (approximately 1 kilometre) is used to represent the topography in the test area. A total of 166 GPS/leveling points in three different networks were used for comparison with the gravimetric geoid in the rough area. The distribution of the GPS points can be seen in Figure 2.

4 Geoid modeling development

Terrain corrections (c_P) were calculated by FFT for each gravity point from the GTOPO30 DEM using the Tc2DFTPL program (Li, 1993). A third order term of a mass prism model was used. The indirect effect on gravity due to Helmert's second method of condensation was considered before applying Stokes's formula and its computation on the geoid should be considered to, at least, second order term.

The maximum indirect effects are correlated with the topography. The statistics of the topographic data and statistics of terrain corrections in gravity stations can be seen in Table 1.

Table 1. Statistics of the GTOPO30. Unit:[m] and c_P. Unit:[mGal]

	Max	Min	Mean	$\pm\sigma$
GTOPO30	6795	0	1399	1465
c_P	42.13	0.20	2.55	2.70

The TC program was used to compute the RTM effects and the topographic-isostatic effects on both gravity and geoid (Forsberg, 1984) and this program was modified to compute the Rudzki inversion method (Bajracharya et al., 2001). The statistics of the gravity anomalies calculated with the four topographic gravity reductions are presented in Table 2.

Table 2. Statistics of the gravity anomalies calculated with the four topographic reductions. Unit [mGal]

	Gravity anomalies	Max	Min	Mean	±σ
Faye (Helmert)	$\Delta g_{FA} + c_P$	258.63	-139.06	30.53	54.95
	$\Delta g_{FA} + c_P - \Delta g_{EGM96}$	155.62	-180.00	-20.62	36.38
Rudzki	$\Delta g_{FA} - \Delta g_{RU}$	255.70	-121.55	41.77	50.87
	$\Delta g_{FA} - \Delta g_{RU} - \Delta g_{EGM96}$	112.28	-113.09	-9.38	27.22
RTM	$\Delta g_{FA} - \Delta g_{RTM}$	285.45	-71.12	64.79	38.66
	$\Delta g_{FA} - \Delta g_{RTM} - \Delta g_{EGM96}$	166.00	-128.37	14.04	37.58
Airy-Heiskanen (AH)	$\Delta g_{FA} - \Delta g_{AH}$	185.01	-74.58	42.74	51.65
	$\Delta g_{FA} - \Delta g_{AH} - \Delta g_{EGM96}$	144.75	-179.69	-8.41	57.57

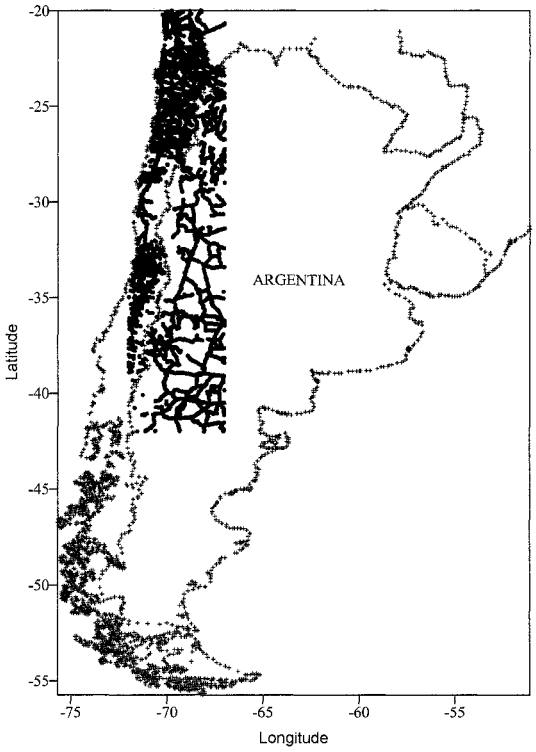


Fig. 1: Distribution of gravity stations in the area under study.

In order to get a good agreement of the gravimetric geoid with respect to the GPS/leveling-derived geoid, the systematic datum differences between the gravimetric geoid and the GPS/leveling data and the long wavelength errors of the geoid were removed by a four-parameter transformation model (Heiskanen and Moritz, 1967). The statistics of the absolute differences between the GPS/leveling derived-geoid and the gravimetric geoids computed using different methods of handling the topography. are summarized in Table 3. The numbers in parentheses refer to the results after the least-squares fitting of the four-parameter transformation model has been applied to the original differences. Before applying the four-parameter transformation model, two GPS on benchmark points with large gross error in either the GPS or the leveling data were removed.

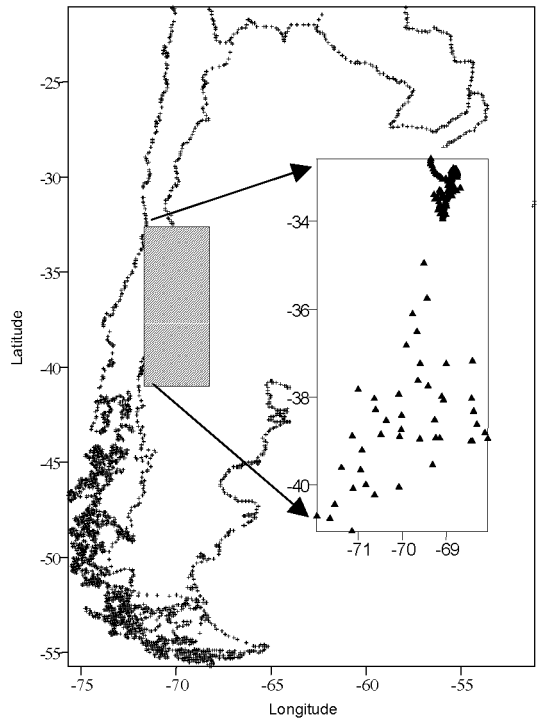


Fig.2: Distribution of GPS/leveling points in the area under study

Table 3. Statistics of different gravimetric geoid solutions (Values in the parentheses are after fit) Unit: [m]

	Max	Min	Mean	±σ
Rudzki	2.86	-1.64	0.44	0.68
	(1.92)	(-1.64)	(0.00)	(0.44)
Helmert	3.37	-1.49	0.71	0.77
	(2.11)	(-1.63)	(0.00)	(0.54)
RTM	2.32	-1.66	1.17	1.10
	(1.09)	(-1.95)	(0.00)	(0.51)
Airy-Heiskanen	-0.81	-8.83	-5.99	1.84
	(-1.75)	(3.43)	(0.00)	(0.61)

5 Conclusions

Four different gravity reduction methods have been presented. They treat the topography in a very different way. Helmert's second method of condensation and the RTM method are the most used reduction techniques for the determination of a gravimetric geoid. In this paper, AH reduction as well as the Rudzki inversion method were also applied. The last one is not used very often, although it has the advantage of no indirect effects.

The gravimetric geoid computed with the Rudzki inversion method gave better results compared with the GPS/leveling-derived geoid before and after fit and was the only method that improved the gravimetric geoid compared to the EGM96 results.

The gravimetric data needs to be improved in the area of the Andes in order to see further improvements in the geoid.

References

Bajracharya, S., Kotsakis C., Sideris M. G., 2001. Geoid Determination Using Different Gravity Reduction Techniques. Presented in IAG meeting, Budapest.

Forsberg, R., 1984. A study of terrain reductions, density anomalies and geophysical inversion methods in gravity field modeling. Report No. 355, Department of Geodetic Science and Surveying, The Ohio State University, Colombus, Ohio, USA.

GTOPO30,1996.
(http://edcdaac.usgs.gov/gtopo30/gtopo30.html).

.Heiskanen, W.A. and Moritz, H., 1967. Physical Geodesy. Freeman and Company, San Francisco.

Lemoine, F. G., Kenyon, S. C., Factim, J. K., Trimmer, R. G., Pavlis, N. K., Chinn, D. S., Cox, C. M., Klosko, S. M., Luthcke, S. B., Torrence, M. H., Wang, Y. M., Williamson, R. G., Pavlis, E. C., Rapp, H. and Olson, T. R., 1998. The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM 96. Pub. Goddard Space Flight Center.

Li, Y., 1993. HFTGVBP Software package for the solution of GVBP by means of fast Hartley/Fourier Transform. TOPOGEOP Software packages to evaluate the TOPOgraphic effects on GEOdetic /GEOPhysical Observation. Department of Geomatics Engineering, The University of Calgary.

Sideris, M.G., She, B.B., 1995. A new high-resolution geoid for Canada and part of U.S. by the 1D-FFT method. Bull Geod 69, 92-108.

Strang van Hees G., (1990). Stokes' formula using fast Fourier techniques. Manuscripta Geodaetica, Vol. 15, pp. 235-239.

Torge, W., 1989. Gravimetry, Walter de Gruyter, Berlin.

Wichiencharoen, C., 1982. The indirect effects on the computation of geoid undulations. OSU Report. No 336, Department of Geodetic Science and Suveying, The Ohio State University, Columbus, Ohio, USA.

On the Numerical Solution of Parameters of the Least Squares Modification of Stokes' Formula

Artu Ellmann
Royal Institute of Technology, Department of Infrastructure, Geodesy Group, SE-100 44 Stockholm, Sweden

Abstract. In regional gravimetric geoid determination, it has become customary to utilize the modified Stokes formula, which combines local terrestrial data with a global geopotential model (GGM). A modification method, proposed by L.E. Sjöberg in 1984 (with later developments), allows least squares minimization of the influence of any error source in geoid modelling. In this approach, depending on the local gravity data quality, the chosen radius of integration, and the characteristics of the used GGM, the modification parameters s_n vary. New satellite gravity missions are expected to improve significantly the accuracy of geopotential models. Of particular interest of this study is to evaluate the impact of future (i.e. post-GOCE) geopotential coefficients. A set of least squares modification parameters is determined from the system of linear equations, aiming at minimizing the global mean square error of geoid estimator. Some difficulties may be encountered when practically computing the modification parameters. In particular, for certain parameters the design matrix suffers from numerical ill-conditioning. Importantly, Tikhonov regularization is satisfactory in providing a solution for the modification parameters. Numerical results are presented to illustrate the applicability of the obtained parameters in geoid modelling by comparing with GPS-levelling data.

Keywords: Stokes' formula, modification, regularization

1 Introduction

The modification of Stokes' formula, originating with Molodensky et al. (1962), aims to reduce the truncation error that occurs when the area of Stokesian integration is limited to a spherical cap. Over recent decades several modification approaches have been published in the literature, which are summarized and cited, e.g. in Featherstone et al. (1998), Sjöberg and Hunegnaw (2000), among others. It should be noted that many of those studies aim at reducing the truncation error only. It is assumed thus, that the errors stemming from the geopotential coefficients and measured gravity anomalies are negligible, an assumption which is obviously not true. With access to recent high-degree GGM-s, the truncation error loses some of its significance, and the GGM errors become more important. One should also consider the erroneous terrestrial data within the integration area. In contrast, the modification methods proposed by Sjöberg (1984, 1991 and 2003) allow least squares (LS) minimization of any known or estimated, random or systematic errors in geoid modelling.

We begin with a brief review of the LS modification theory and relevant expressions. Since for certain LS methods some numerical difficulties may be encountered, this study aims of contributing the important computational aspects of the parameters s_n. Apparently, there is a need to determine the parameters from an ill-conditioned linear system. Hence, a special regularization procedure must be undertaken to stabilize the numerical solution. The third paragraph is dedicated to the principles of Tikhonov regularization, which then will be applied in the numerical studies. To illustrate the applicability of Tikhonov regularization for the LS parameter determination, the results of geoid modelling in a test area are presented. Additionally, the impact of future (i.e. post-GOCE) GGM coefficients for LS modification is studied. A short conclusion follows.

2 Least Squares Modification of Stokes' Formula

Stokes' formula enables the determination of the separation N between the geoid and geocentric reference ellipsoid from the global coverage of the gravity anomalies Δg:

$$N = \frac{R}{4\pi\gamma} \iint_{\sigma} S(\psi)\Delta g d\sigma \,, \tag{1}$$

where R is the mean Earth's radius, γ is normal gravity on the reference ellipsoid, ψ the geocentric angle, the Stokes function, $S(\psi)$, can be found in most of geodetic textbooks, and $d\sigma$ is an infinitesimal surface element of the unit sphere σ. Rigorously, Eq. (1) should also comprise some correction terms, e.g. accounting for Earth's

ellipticity, the contribution of topographic and atmospheric masses, etc. It is assumed, however, that such effects can be accounted separately, and are thus omitted in this paper. The double integral in Eq. (1) has to be evaluated over the whole Earth. However, in practice the area of integration is usually limited to a spherical cap around the computation point. The truncation error of the remote zone can be reduced by introducing a modification of Stokes' formula, which combines the terrestrial gravity anomalies and a GGM. Assuming a cap of integration σ_0 around the computation point, a general estimator $\tilde{N}$ of the geoidal height is provided by two sets of parameters (s_n and b_n), and gravity anomalies $\Delta\hat{g}$, inserted into the modified Stokes formula

$$\tilde{N} = \frac{R}{4\pi\gamma}\iint_{\sigma_0} S^L(\psi)\Delta\hat{g}\,d\sigma + \frac{R}{2\gamma}\sum_{n=2}^{M} b_n \Delta\hat{g}_n \quad , \qquad (2)$$

where the anomaly degree estimate, $\Delta\hat{g}_n$, can be calculated from a GGM, and the modified Stokes function is expressed by selected parameters s_k and in terms of *Legendre* polynomials $P_k(\cos\psi)$

$$S^L(\psi) = S(\psi) - \sum_{k=0}^{L}\frac{2k+1}{2}s_k P_k(\cos\psi) \, . \qquad (3)$$

The upper modification limit M is arbitrary (often chosen by the upper limit of the available GGM), and generally not equal to L, see Eq. (3).

First of all, the LS estimation of the parameters aims to minimize the expected global mean square error of the estimator, Eq. (2). Regarding the spectral form of the geoidal undulation N the global error of $\tilde{N}$ can be written:

$$m_{\tilde{N}}^2 = E\left\{\frac{1}{4\pi}\iint_{\sigma}(\tilde{N}-N)^2 d\sigma\right\} = c^2\sum_{n=2}^{M} b_n^2 dc_n + \\ + c^2\sum_{n=2}^{\infty}\left[\left(b_n^* - Q_n^L - s_n^*\right)^2 c_n + \left(\frac{2}{n-1} - Q_n^L - s_n^*\right)^2 \sigma_n^2\right] \qquad (4)$$

where $c = R/2\gamma$, and

$$b_n^* = \begin{cases} b_n, & \text{if } 2 \le n \le M \\ 0 & \text{otherwise} \end{cases} \; ; \; s_n^* = \begin{cases} s_n, & \text{if } 2 \le n \le L \\ 0 & \text{otherwise} \end{cases} \, . \qquad (5)$$

The truncation coefficients Q_n^L are calculated

$$Q_n^L = Q_n - \sum_{k=0}^{L} E_{nk} s_k \quad , \qquad (6)$$

whereas the coefficients Q_n and $E_{nk} = (k+0.5e_{nk})$ are the functions of the truncation radius of the integration cap. It is notable, that the mean degree variances of signal and noise are applied in Eq. (4). The gravity anomaly variances are denoted by c_n, whereas the terrestrially measured and GGM-derived anomaly error degree variances are denoted by σ_n^2 and dc_n, respectively. Note also that current high degree GGM are determined from a combination of satellite data and terrestrial gravity data. This combination implies that the two data sets in Eq. (2) are frequently correlated. Since the main scope of present study is determination of the LS parameters this correlation is not considered here. All errors are assumed to be random and with expectations zero. For the full theory and complete expressions (accounting also for systematic errors and correlated data sets) of the LS modification the interested reader is referred to Sjöberg (1984 and 1991). Note that the methods have some similarities to Wenzel's (1983) least squares spectral combination method, which also reduces the errors stemming from the terrestrial data and the GGM.

Eventually, by selecting the parameters s_n in Eq. (4) in a suitable way (see Table 1) it should be possible to minimize $m_{\tilde{N}}^2$. For this, Eq. (4) is differentiated with respect to s_n, i.e. $\partial m_{\tilde{N}}^2 / \partial s_n$. The obtained formula is then equated to zero, and the modification parameters s_n are solved from the linear system of equations:

$$\sum_{r=2}^{L} a_{kr} \cdot s_k = h_k \, , \quad k, r = 2,3,\ldots L \, , \qquad (7)$$

where a_{kr} and h_k are modification coefficients, which can be expressed via Q_n, E_{nk}, c_n, dc_n and σ_n^2.

In this paper the focus is given to the biased, unbiased and optimum method of LS modification, which basically deviate from each other in the choice of parameters b_n. Secondly, in the two latter methods more parameters s_n are allowed than degrees of potential coefficients available. After LS

Table 1. Main parameters of the least squares modification of Stokes' formula.

1 Biased ($L = M$)	$b_n = s_n$	$(q_n)_1 = c_n$
2 Unbiased ($L \ge M$)	$b_n = s_n + Q_n^L$	$(q_n)_2 = \begin{cases} dc_n, & \text{if } 2 \le n \le M \\ c_n & \text{otherwise} \end{cases}$
3 Optimum ($L \ge M$)	$b_n = \frac{(s_n + Q_n^L)c_n}{c_n + dc_n}$	$(q_n)_3 = \begin{cases} \frac{c_n dc_n}{c_n + dc_n}, & \text{if } 2 \le n \le M \\ c_n & \text{otherwise} \end{cases}$

parameters s_n are defined, the expected global error, $m_{\hat{N}}^2$, can be estimated from

$$m_{\hat{N}}^2 = f_i - c^2 \hat{\mathbf{s}}_i^T \mathbf{h}_i \;\; ; \;\; i = 1,2,3 \;\; , \tag{8}$$

where $\hat{\mathbf{s}}$ and $\mathbf{h}$ are the vectors of coefficients, and the index i refers to the respective LS method, and

$$f_i = c^2 \sum_{n=2}^{\infty} \left[\left(\frac{2}{n-1} - Q_n \right)^2 \sigma_n^2 + Q_n^2 \left(q_n \right)_i \right] , \tag{9}$$

lastly, $(q_n)_i$ is presented in Table 1.

Some difficulties may be encountered when practically computing the LS modification parameters. The system of Eq. (7) can be defined as $\mathbf{Ax} = \mathbf{b}$, with vector $\mathbf{b}$ consisting of the modification coefficients h_k, the design matrix $\mathbf{A}$ consisting of the coefficients a_{kr}, and $\mathbf{x}$ is the vector of unknown modification parameters s_n. For some LS methods the matrix $\mathbf{A}$ may become ill-conditioned, and tedious manipulations have to be taken for inverting $\mathbf{A}$ and to find the required LS parameters.

3 Regularization of the Ill-Conditioned Linear Systems

A problem is improperly posed if it does not meet at least one of the following three requirements: existence, uniqueness, or stability of the solution. A useful indicator of non-stability or ill-condition is the condition number of the matrix $\mathbf{A}$, which is defined as the ratio between the largest and the smallest singular values: $cond(\mathbf{A}) = \mu_{max}/\mu_{min}$. Obviously, matrices with $1/cond(\mathbf{A})$ reaching or exceeding the computational precision of the computers are very ill-conditioned. This also pertains, that noisy elements and rounding errors may be greatly amplified during the inversion process of $\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$, which may lead to a physically meaningless solution. A way out of this problem is to incorporate some kind of regularization in the solution procedure.

For solving the LS parameters Tikhonov regularization will be applied in this study. This method is among other useful regularization methods explicitly described in the literature; see e.g. Hansen (1998). Regularization usually aims to minimize the norm of difference between two vectors - $\min\|\mathbf{Ax}-\mathbf{b}\|$. The Tikhonov regularization scheme can be expressed as

$$\min\left\{ \left\| \mathbf{Ax} - \mathbf{b} \right\|_2^2 + \lambda^2 \left\| \mathbf{Lx} \right\|_2^2 \right\} , \tag{10}$$

where λ is the regularization parameter and $\mathbf{L}$ is a diagonal weighting matrix (in this study the standard $\mathbf{L}$ is assumed, i.e. the identity matrix $\mathbf{L}=\mathbf{I}$). The Tikhonov regularized solution, $\mathbf{x}_{reg}$, can be formulated through the normal equation matrix

$$\mathbf{x}_{reg} = \left(\mathbf{A}^T\mathbf{A} + \lambda^2 \mathbf{L}^T\mathbf{L} \right)^{-1} \mathbf{A}^T\mathbf{b} \;\; . \tag{11}$$

It is appreciated that this solution can also be computed by means of the singular value decomposition (SVD) of $\mathbf{A}$

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^{L} \mathbf{u}_i \mu_i \mathbf{v}_i^T \;\; , \tag{12}$$

where $\mathbf{U}$ and $\mathbf{V}$ are matrices with orthonormal columns ($\mathbf{u}_i$ and $\mathbf{v}_i$, respectively) satisfying the condition $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$, and the diagonal matrix Σ has non-negative singular values of $\mathbf{A}$ appearing in non-increasing order such that $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_L$ (Bjerhammar 1973, Chap. 27.9). Finally, the regularized solution is expressed as

$$\mathbf{x}_{reg} = \sum_{i=1}^{L} f_i \frac{\mathbf{u}_i^T\mathbf{b}}{\mu_i} \mathbf{v}_i \;\; . \tag{13}$$

Inspection of the elements in Eq. (13) gives good insight into the problem. Intuitively, the norm $\left\| \mathbf{x}_{reg} \right\|$ becomes very large due to a small μ_i in denominator, and without any regularization the solution will therefore be dominated by the last SVD components. In Eq. (13) the numbers f_i are filter factors for the chosen regularization method. The filter factors must have the important property that as μ_i decreases, the corresponding f_i tends to zero in such a way that the contributions from the smaller μ_i are effectively filtered out. Particularly, the filter factor for Tikhonov regularization of a symmetric matrix (recall that $a_{kr}=a_{rk}$ in LS modification) could be defined as (cf. Hansen 1998, p 107)

$$f_i = \frac{\mu_i}{\left(\mu_i + \lambda \right)} \;\; . \tag{14}$$

No regularization method is complete without a method for choosing the regularization parameter λ. Perhaps the most convenient graphical tool for defining the Tikhonov regularization parameter is the so-called L-curve, which is a plot for all valid parameters λ of the regularized solution norm $\left\| \mathbf{Lx}_{reg} \right\|$ versus the residual norm $\left\| \mathbf{Ax}_{reg} - \mathbf{b} \right\|$. When the L-curve is plotted in log-log scale (for an

ill-conditioned matrix singular values span many orders of magnitude), it has very often a characteristic L-shaped appearance (hence its name) with a distinct corner separating the vertical and the horizontal parts of this curve. The optimal regularization parameter is often determined as L-curve corner-point, which corresponds to a solution in which the regularization and perturbation errors are balanced. The exact definition of the corner point is not a trivial problem and different methods may be applied (e.g. see Hansen 1998).

4 Numerical Investigations

4.1 Initial Data

In recent computations of geoid models at the Royal Institute of Technology the biased LS method, which gives us a well-conditioned solution for the modification parameters s_n, were used. Most likely the other LS methods have not been applied for regional geoid modelling, yet. First of all, this could be explained by occurrence of ill-conditioning in parameter determination. Nevertheless, Sjöberg and Hunegnaw (2000) reported on solving the unbiased LS parameters for two special cases when $L=M=20$ and $L=M=180$, and the terrestrial anomaly errors were switched to zero. It is acknowledged that the accuracy estimation of the terrestrial data is rather delicate task. Recall that the data is usually collected during several decades with different methods and specifications. However, it is better to use a coarse error model, than to assume that data are without errors. Some idea of the data accuracy can be obtained empirically when comparing the gridded anomaly values with known control points.

For estimation of the terrestrial gravity anomaly degree variance σ_n^2 some assumptions for the covariance function $C(\psi)$, the error variance $C(0)$ (the value of the covariance function at $\psi = 0$) and the correlation length (to the point where $C(\psi)$ is decreased to half of $C(0)$, see Moritz 1980, p. 174) need to be adopted. In this study the closed expression for $C(\psi)$ in Sjöberg (1986, eq. 7.2) and the correlation length as of 0.1° are applied. The terrestrial anomaly error variance value $C(0)$ vary from 9 to 16 mGal^2. Note that generally the variations in chosen $C(0)$ values have only minor impact to the resulting least squares $S^L(\psi)$.

Three values of integration cap radii, ψ_0, with spherical distance 2°, 3° and 5° (corresponding to 220, 330 and 550 km, respectively), were tested. The coefficients Q_n and e_{nk} in Eq. (6) were expanded to maximum degree 2000 (approx. ∞).

Gravity anomalies and their error degree variances (c_n and dc_n) were estimated from the Earth Geopotential Model EGM96 (Lemoine et al. 1998), which coefficients are complete to maximum degree and order 360. Higher degree c_n could be generated synthetically to conform with the predicted spectral characteristics of the Earths gravity field (e.g. following Kaula's rule, although its validity in very high degrees may be questioned).

The modification degree M is in this study equated to the maximum degree of the EGM96, i.e. $M = 360$. This choice is supported by the circumstance, that if a GGM will be truncated, we lose the valuable high-degree information (since the contribution of truncated harmonics cannot be reflected in lowest terms of expansion). The coefficients a_{kr} and h_k are derived from Eq. (4), when applying differentiation $\partial m_N^2 / \partial s_n = 0$.

4.2 Regularization and LS Parameters

Inspection of the outcome of the SVD for each coefficient matrix **A** is useful. Generally, and independent on the chosen combination of input values, the condition number for the biased LS method is approximately 10^3, when for the two other LS methods it ranges to about 10^{18}. Standard double precision for most software packages is 10^{-16} Hence, for reducing the rounding errors, the extended computational precision may be suggested. The SVD elements of unbiased and optimum methods range over 18 orders of magnitude, which yield an extremely ill-conditioned system of linear equations. In contrast, the system for the biased LS method is relatively well-conditioned, which allows us to find the unknowns s_n as the product of outright inverse of matrix **A** and a vector of h_k coefficients. For the ill-conditioned systems the straightforward inverse of **A** leads to parameters s_n, which oscillate within the range of $\pm 10^5$ and are thus meaningless. In these cases the Tikhonov regularization parameter λ is defined by the maximum curvature of the L-curve, which is determined from classical differentiation formulas. Importantly, Tikhonov regularization, Eq. (13), and SVD, reveal to be satisfactory for obtaining applicable parameters s_n for any combination of input data sets in unbiased and optimum methods.

As a next step, each set of s_n parameters, coefficients Q_n and E_{nk} were inserted into Eq. (6), which resulted in respective sets of Q_n^L. These coefficients were thereafter used for computing the parameters b_n (see Table 1). An interesting feature is, that for the same initial conditions (i.e. $C(0)$, M, L, and ψ_0) the resulting b_n sets of the three LS

methods (biased, unbiased, optimum) are numerically very similar. The variations between different LS methods are insignificant, which allowed us to denote the LS solutions with a single **bold** line in Figure 1.

In contrast, the Molodensky coefficients $b_n = Q_n$ are *a priori* fixed pertaining thus the truncation error as the only error source to be reduced. The parameters b_n for the modification cases start from $n = 2$. The first LS parameter, b_2, is slightly less than 2. The following b_n values are quickly decreasing. After reaching zero these start to converge and oscillate around the zero axes. The convergence of the Molodensky modification coefficients is much slower (Figure 1). Furthermore, the typical behaviour of $S^L(\psi)$ {by Eq. (3)}, with different sets of least squares parameters s_n, across the integration cap was investigated. Importantly, any LS parameter set "forces" the modified Stokes' function to converge to zero at the edge of the integration cap, see Figure 2. Molodensky modification applies the original Stokes function $S(\psi)$ {i.e. all s_n are set to zero in Eq. (6)}, also portrayed in Figure 2.

4.3 Results and Accuracy of LS Methods

The quality of the obtained LS parameter sets could be evaluated by the global error, Eq. (8). Certainly, variations in any of the input coefficients and conditions can slightly improve or worsen this error. However, the magnitude of the global error for any computed LS parameter set s_n is around 0.12-0.15 m. Formally, the global mean error was smallest for the optimum LS method. However, different results of geoid modelling can be expected in different regions.

The more reliable assessment of a particular modification method is the comparison with GPS and levelling data. The inter-comparison of a geoid

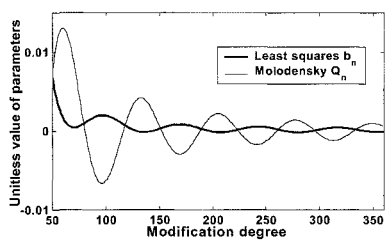


Fig. 1 Least squares modification parameters b_n (EGM96 as the reference model, modification degree $L = M = 360$, variance of terrestrial data $C(0) = 9$ mGal^2, integration cap size $\psi_0 = 5°$) versus the Molodensky coefficients. Note that due to scaling the first 50 degrees are not displayed.

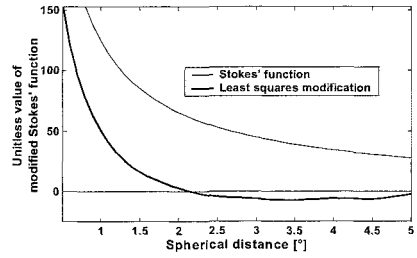


Fig. 2 Behaviour of the modified Stokes function $S^L(\psi)$, {LS modification, EGM96 as the reference model, modification degree $L = M = 360$, cap size $\psi_0 = 5°$, variance of terrestrial data $C(0) = 9$ mGal^2} versus original Stokes' function $S(\psi)$. Note that due to scaling the values for the distance $\psi \leq 0.5°$ are not displayed.

model, GPS-derived and spirit-levelled heights at discrete points gives a reasonable indication of the gravimetric geoid model accuracy. The LS modification of Stokes' formula was applied for computing geoid models for a test area (approx. 83500 km^2), which includes the whole Estonia. The gravity data outside of the borders of the target area are involved as well. The precision of the free-air anomaly grid in Estonia is estimated as 2.4 mGal.

We apply biased, unbiased and optimum LS methods. The respective parameters b_n and s_n, are calculated from the following initial conditions: (1) EGM96 as reference model, (2) modification degrees $L = M = 360$, (3) terrestrial anomaly variance $C(0) = 9$ mGal^2 (the correlation length as of 0.1° assumed), (4) integration cap size $\psi_0 = 2°$.

The computed geoid models (grid size 1.5′x3′, corresponding to 2.7x2.7 km^2) were validated by 26 nationwide Estonian geodetic points. In these points the combined error of GPS-derived and spirit levelled heights cannot be larger than 2-3 cm. These 26 points form the "geometric geoid model" which may be considered as the practical realisation of the height system. The set of 26 GPS-levelling points was then used for the definition of the four transformation parameters among the pairs of geoid models. A root mean square error of post-fitting residuals at GPS-levelling points within 3-4 cm was achieved for any of the tested geoid models. It could be concluded, that the accuracy of the three LS methods of modification of Stokes' formula is of the same level as is the accuracy of the used control points. The computations described above were repeated with some other (and more recent) GGM. Tests included cases with M lower than 360 and $L > M$ as well. The resulting accuracy (both for global error and post-fitting residuals) was found almost similar as for the EGM96 applications.

4.4 Improvements by GOCE

Major expectations for the improved GGM are related to the already launched (or planned) satellite gravity missions. Of particular interest is GOCE, which will provide unprecedented accuracy for geopotential coefficients in the global scale and up to degree 270 (corresponding to the spatial resolution of 65 km). However, for many practical applications a more refined regional geoid model may be necessary. Eventually, new terrestrial or airborne data to be captured also need to be included into geoid modelling. Thus, the modification of Stokes formula does not lose its importance.

To evaluate the impact of improved accuracy for LS parameters the simulated error degree variances (cf. Rummel et al. 2000) dc_n could be utilized. Other coefficients $\{C(0), M, L,$ and $\psi_0\}$ remain the same (including also c_n, which were computed from the existing GGM-s, though the improved gravity anomaly from GOCE could be expected). New calculations of the LS modification parameters were produced. Again, for the ill-conditioned unbiased and optimum LS methods Tikhonov regularization with optimum parameter as L-curve corner-point was applied. The new global mean errors for the three LS methods are in the range of 0.03-0.04 m, which is 3-4 times better than is obtained with the accuracy characteristics of existing GGM-s (see Section 4.3). This agrees with common expectations that the practical geoid modelling in post-GOCE era may provide significantly improved results than up-to-date GGM-s.

5 Conclusion

This study concerned the practical solution of the least squares parameters of the modified Stokes formula. The least squares modification allows to minimize any known or estimated error source. In this approach, depending on the local gravity data quality, the chosen radius of integration, and the characteristics of the used GGM, the modification parameters s_n vary. In particular, the parameters of unbiased and optimum LS modifications need to be defined from an ill-conditioned system. The validity of LS modification is verified through the application of Tikhonov regularization with an optimum parameter as L-curve corner-point. Importantly, employing the singular value decomposition for regularization is very transparent and easy to implement in practical computations. Numerical tests support the suitability of LS parameters for a regional geoid modelling. The impact of future, i.e. post-GOCE, geopotential models was also evaluated. Obtained results demonstrate, that LS modification of Stokes' formula is well suited for computing the high-resolution and accurate geoid models.

Acknowledgements. The author is grateful to Prof. L. Sjöberg for a constructive review of an early version of the manuscript. An anonymous reviewer is thanked for valuable comments. This study was supported by the Kami Research Foundation, and the Lindstrands Foundation awarded the IUGG 2003 travel grant. This support is cordially acknowledged.

References

Bjerhammar A (1973) *Theory of Errors and Generalized Matrix Inverses.* Elsevier, Amsterdam.

Featherstone WE, Evans JD, Olliver JG (1998) A Meissl-modified Vanicek and Kleusberg kernel to reduce the truncation error in gravimetric geoid determination. *Journal of Geodesy,* 72, 154-160.

Hansen PC (1998) *Rank-deficient and Discrete Ill-Posed Problems: numerical aspects of linear inversion.* SIAM, Philadelphia.

Lemoine FG, Kenyon SC, Factor JK, Trimmer RG, Pavlis NK, Chinn DS, Cox CM, Klosko SM, Luthcke SB, Torrence MH, Wang YM, Williamson RG, Pavlis EG, Rapp RH, Olson TR (1998) *The development of the joint NASA GSFC and NIMA Geopotential Model EGM96.* NASA/TP-1998-206861.

Molodensky MS, Eremeev VF, Yurkina MI (1962) *Methods for study of the external gravitational field and figure of the Earth.* Israel program for Scientific translations, Jerusalem.

Moritz H (1980) *Advanced physical geodesy.* Herbert Wichmann, Karlsruhe/Abacus Press, Tunbridge Wells.

Rummel R, Müller J, Oberndorfer H, Sneeuw N (2000) Satellite Gravity Gradiometry with GOCE. In: Rummel R, Drewes H, Bosch W, Hornik H (eds.) *Towards an Integrated Global Geodetic Observing System.* IAG Symp. Series 120, 66-72, Springer Verlag.

Sjöberg LE (1984) *Least squares modification of Stokes and Vening-Meinesz formulas by accounting for errors of truncation, potential coefficients and gravity data.* The Dept. of Geodesy report No. 27, University of Uppsala.

Sjöberg LE (1986) Comparison of some methods of modifying Stokes formula. *Bollettino di Geodesia e Scienze Affini,* 46, 229-248.

Sjöberg LE (1991) Refined least-squares modification of Stokes formula. *Manuscripta Geodaetica,* 16, 367-375.

Sjöberg LE (2003) A General Model of Modifying Stokes' formula and Its Least-Squares Solution. Accepted by *Journal of Geodesy.*

Sjöberg LE, Hunegnaw A (2000) Some modifications of Stokes formula that account for truncation and potential coefficient errors. *Journal of Geodesy,* 74, 232-238.

Wenzel H-G (1983) Geoid computation by least squares spectral combination using integral kernels. *Proceedings IAG General Meeting,* Tokyo, pp. 438-453, Springer Verlag.

A Numerical Comparison of Different Ellipsoidal Corrections to Stokes' Formula

Artu Ellmann
Royal Institute of Technology, Department of Infrastructure, Geodesy Group, SE-100 44 Stockholm, Sweden

Abstract. This paper reviews the new method (recently derived by L.E. Sjöberg) of the ellipsoidal correction for Stokes' formula. Importantly, the correction can also be expressed in a series of spherical harmonics of the disturbing potential, which is a considerable computational advantage. In order to assess the applicability of this approach, it is numerically compared with two other methods. The results reveal that the new method practically coincides with one of the earlier methods, although their derivational approaches and resulting expressions are considerably different. The new method is adapted for the modified Stokes formula, which combines regional terrestrial gravity data with a global geopotential model. The magnitude of the ellipsoidal correction in the modified Stokes formula does not exceed cm level, globally.

Keywords: geoid, ellipsoidal correction, Stokes' formula, least squares modification.

1 Introduction

Geoid determination by Stokes' formula rigorously holds only on a spherical boundary, yielding that the gravity anomalies Δg must be referred to the mean Earth sphere (MES, with radius R). In practice, however, observed on the Earth's surface, anomalies Δg are commonly reduced to mean sea level (MSL). This implies that the right side of the well-known boundary condition

$$\Delta g = -\frac{\partial T}{\partial r} - \frac{2}{r}T \quad , \tag{1}$$

(where r is geocentric radius, and T is the disturbing potential) must be upward or downward continued to MES, before Stokes' integral can be applied. The MSL deviates from the reference ellipsoid by ±100 m at most, allowing its approximation by an ellipsoid. The ellipsoidal effect becomes more apparent if the Stokes formula is written in the form

$$N = N^0 + \delta N_{ell} = \frac{R}{4\pi\gamma}\iint_{\sigma} S(\psi)(\Delta g + \delta g)d\sigma \quad , \tag{2}$$

where N is the geoidal height to be determined, ψ the geocentric angle between integration points, the Stokes function $S(\psi)$ can be found in most geodetic textbooks, Δg is the surface gravity anomaly (referred to sea level), δg is the ellipsoidal effect to the anomaly, γ is normal gravity on the reference ellipsoid, and $d\sigma$ is an infinitesimal surface element of the unit sphere σ. Note that Eq. (2) is the sum of the spherical Stokes integral and the ellipsoidal correction δN_{ell} to the geoidal height. It is assumed that other effects (e.g. related to the Earth's topography and atmosphere) in the geoid modelling can be solved separately, and are thus omitted in this paper.

The procedure of transforming the right side of Eq. (1) to the MES, however, is often neglected in practical computations. This yields an error in geoid determination of the order of the Earth's geometrical flattening (approx. 1/300) times the geoidal height, reaching thus several decimetres, globally. Hence, for an accurate geoid model it is important to estimate the ellipsoidal correction.

Different authors have studied the ellipsoidal correction for the original Stokes' formula, see e.g. Molodensky et al. (1962), Moritz (1980), Martinec and Grafarend (1997), Fei and Sideris (2000), and references therein. For the derivation of the ellipsoidal correction different boundary conditions are posed, where the disturbing potential is commonly expressed as a series expansion of eccentricity e^2. The solutions differ in various respects, preventing their direct analytical comparison. Two recent studies by Huang et al. (2003), and Heck and Seitz (2003), tackle exhaustively the above issues. Since their results were published shortly after the compilation of the previous version of the manuscript, this prompted to some revision of our paper.

We begin with a brief review of two ellipsoidal correction methods: Martinec and Grafarend (1997), and, Fei and Sideris (2000). A new integral solution for the ellipsoidal correction was derived by Sjöberg (2003a). Importantly, this solution can be developed into a spherical harmonic series. This study aims to contribute with the numerical verification of the new ellipsoidal correction.

It should also be noted, that most published methods concern the case of the original Stokes formula only. In practice, the area of integration is

usually limited to a spherical cap around the computation point. Nowadays, the estimation of a regional geoid model typically relies on the modified Stokes formula, which combines the terrestrial anomaly Δg and a global geopotential model. In contrast to the original formula, the ellipsoidal corrections for the modified Stokes formula, have received less attention in the geodetic literature, and are often neglected in the practical computations. Sjöberg's method, adapted to account the ellipsoidal correction for the modified Stokes formula, is reviewed in the third section. Finally, numerical results are presented to illustrate the range of the global distribution of the ellipsoidal correction for each of the reviewed methods.

2 Ellipsoidal Corrections to Stokes' Formula

2.1 Martinec and Grafarend (1997) method

Martinec and Grafarend (1997) formulated the boundary value problem of physical geodesy on the surface of the reference ellipsoid by means of ellipsoidal harmonics (see Heiskanen and Moritz 1967, pp. 27-28). The anomalies thus need to be referred to the surface of this ellipsoid, yielding a small correction to observed Δg. The authors presume the ellipsoidal approximation of the geoid reflecting reality much better than commonly applied spherical approximations. Note that in the definition of the geoidal height N^{MG} radius R is replaced by semi-minor axis b {cf. Eq. (2)}

$$N^{MG} = \frac{b}{4\pi\gamma}\iint_{\Omega} S(\psi)\Delta g d\Omega + \delta N_{ell}^{MG} \quad , \qquad (3)$$

where $d\Omega$ is an infinitesimal surface element of the ellipsoid. The Martinec-Grafarend ellipsoidal correction is expressed in the form

$$\delta N_{ell}^{MG} = e^2 \frac{b}{4\pi\gamma}\iint_{\Omega} S^E(\vartheta,\psi,\alpha)\Delta g d\Omega \quad , \qquad (4)$$

where the 'ellipsoidal Stokes function' S^E is a function of ϑ (the reduced ellipsoidal co-latitude, for its definition see Heiskanen and Moritz 1967, pp. 39-40), azimuth α and spherical distance ψ between the integration points, respectively. The closed spatial representation of S^E can be found in Martinec and Grafarend (ibid., eq. 50). Recall that ellipsoidal coordinates are applied in this method, and Δg differs slightly from the one in Eq. (2).

2.2 Fei and Sideris (2000) method

The geoidal height in the Fei and Sideris approach is computed by adding the ellipsoidal correction to the approximate geoidal height N^0 {cf. Eq. (2)}

$$N^{FS} = \frac{R}{4\pi\gamma}\iint_{\sigma} S(\psi)\Delta g d\sigma + \delta N_{ell}^{FS} = N^0 + \delta N_{ell}^{FS} \quad , \qquad (5)$$

where the Fei-Sideris correction is presented as a function of N^0 and could be decomposed as

$$\delta N_{ell}^{FS} = e^2 N_{11} + \frac{e^2}{4\pi}\iint_{\sigma} N^0 f_0(\psi,\theta,\theta')d\sigma \quad , \qquad (6)$$

where the term N_{11} is explicitly represented in Fei and Sideris (ibid., eq. 45), and is a function of the harmonic coefficients of the disturbing potential, N^0, mean gravity, and geocentric co-latitude θ of the computational point. The closed formula of the kernel function $f_0(\psi,\theta,\theta')$ is defined in Fei and Sideris (2000).

Note that for computing δN_{ell}^{FS} there is a need for an auxiliary data set N^0. The authors suggest that for an accurate correction it is sufficient to compute the far-zone N^0 from an existing geopotential model, though, higher resolution N^0 values are required for the near-zone.

2.3 Sjöberg (2003a) method

It is remarkable that the derivations in Fei and Sideris (2000) are initially based on the sphere with radius b, i.e. the sphere that goes through the poles of the Earth ellipsoid. In contrast, the new method for ellipsoidal correction by Sjöberg (2003a) adopts the sphere with radius a (semi-major axis of the Earth ellipsoid) as initial boundary for the derivation and numerical analysis. For theory the interested reader is referred to Sjöberg (ibid.). The final expression for the ellipsoidal correction of Sjöberg's method {cf. Eq. (2)} is

$$\delta N_{ell} = N - N^0 = kN^0 + \frac{r_e - a}{\gamma}\left(\frac{\partial T}{\partial r}\right)_{r=a} + \\ + \frac{a}{4\pi\gamma}\iint_{\sigma} S(\psi)\left[\delta g^1 + (a - r_e)\frac{\partial \Delta g}{\partial r}\bigg|_{r=a}\right] d\sigma \quad , \qquad (7)$$

where r_e is geocentric radius of the computation point on the ellipsoid. The scale factor $k = (a-R)/R$ is needed for integration over the bounding sphere instead of the MES. The validity of this simple term

may be questioned, since according to our present knowledge such a factor is not present in any other papers on the subject. However, we are confident in the necessity of this term. For a proof see Sjöberg {2003a, Chap. 3; also in Sjöberg 2003c}.

The middle term in Eq. (7) transfers T from the bounding sphere to the Earth's ellipsoidal surface. The brackets [ ] within the integral of Eq. (7) link the anomaly to the bounding sphere. This term is obtained from relations in Cruz (1986), where

$$\delta g^1 = e^2 \left(2 - 3\cos^2\theta - \sin\theta\cos\theta \frac{\partial}{\partial\theta} \right) \frac{T}{a} \quad . \qquad (8)$$

However, the form of the ellipsoidal correction to Stokes' formula by Eq. (7), is not useful for a numerical computation, since it includes the disturbing potential and its radial derivatives {see also Eq. (1)}. Moreover, all reviewed methods for ellipsoidal corrections are integral formulas, which suffer from the singularity in the vicinity of the computation point. Since the long wavelength features are expected to dominate in the ellipsoidal correction, it would be practical to replace the global integration by the application of spherical harmonics. In Sjöberg (2003a) the ellipsoidal correction, Eq. (7), was developed into a solution, which employs harmonics of T. The full theory and explicit expressions are presented in Sjöberg (ibid.). The ellipsoidal correction is computed by

$$\delta N_{ell} = \delta N_{0,1} + kN^0 + e^2 \frac{GM}{2a\gamma} \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \frac{Y_{nm}}{n-1} \{ [3 - (n+2) F_{nm}] C_{nm} - (n+1) G_{nm} C_{n-2,m} - (n+7) E_{nm} C_{n+2,m} \} \qquad (9)$$

where the zero- and first-degree contribution, $\delta N_{0,1}$, can be computed by (cf. Sjöberg 2003a, eq. 55b)

$$\delta N_{0,1} = e^2 \frac{GM}{2a\gamma} \sum_{n=0}^{1} (2n+1) \sum_{m=-n}^{n} E_{nm} C_{n+2,m} Y_{nm} \, , \qquad (10)$$

GM is the product of the Earth's mass and the gravitational constant, n and m are the degree and order of harmonic series, the coefficients C_{nm} are the fully normalised harmonic coefficients of the disturbing potential, Y_{nm} is the fully normalized spherical harmonics (cf. Heiskanen and Moritz 1967, p. 31), and the coefficients E_{nm}, F_{nm}, G_{nm} can be derived through relations presented in Martinec (1998). Note that Stokes' formula is blind to the zero- and first-degree terms, thus $C_{00} = C_{1|m|} = 0$. Importantly, the solution by Eq. (9) is convenient for practical computations, since the singularity problem is eliminated and tedious global integration is avoided. For alternative ellipsoidal corrections in spherical harmonic representation see Moritz (1980, Chap. 39), and Heck and Seitz (2003, eq. 59). It is notable that if the term kN^0 in Eq. (9) is excluded, then the correction expressions in the Sjöberg and Heck-Seitz methods for degrees $n \geq 2$ are practically equivalent. At the same time the contribution of zero- and first-degrees is accounted differently, although the numerical discrepancies reach only about 1-2 cm, globally (see Section 4).

3 Ellipsoidal Correction to the Modified Stokes Formula

When working with Stokes' formula, the integral in Eq. (2) has to be evaluated over the whole Earth, which is an impractical requirement. Therefore the area of integration is usually limited to a spherical cap around the computation point. The lack of a global coverage of gravity data can be compensated by a combination of terrestrial gravity with the Earth geopotential model (EGM). Assuming a cap of integration σ_0 around the computation point, the unbiased estimator $\tilde{N}$ of the geoidal height is provided by the parameters s_n, and Δg, inserted into the modified Stokes formula (cf. Sjöberg 1991)

$$\tilde{N} = \frac{R}{4\pi\gamma} \iint_{\sigma_0} S^L(\psi) \Delta g \, d\sigma + \frac{R}{2\gamma} \sum_{n=2}^{M} \left(Q_n^L + s_n \right) \Delta\hat{g}_n \, , \qquad (11)$$

where the anomaly degree estimate, $\Delta\hat{g}_n$ can be calculated from an EGM, the modified truncation coefficients Q_n^L are functions of the limiting radius of integration cap, and the modified Stokes function is expressed by selected parameters s_k and in terms of *Legendre* polynomials $P_k(\cos\psi)$

$$S^L(\psi) = S(\psi) - \sum_{k=0}^{L} \frac{2k+1}{2} s_k P_k(\cos\psi) \, . \qquad (12)$$

The upper modification limit M is arbitrary, and generally not equal to L, see Eq. (12).

Several modification approaches have been published since the pioneering work of Molodensky et al. (1962). This paper focuses only to the least squares modification of Stokes' formula (Sjöberg 1991). In this approach, depending on the local gravity data quality, the chosen radius of integration, and the characteristics of the used EGM, the modification parameters s_n vary. For the details of the determination of the parameters s_n the

interested reader is referred to Ellmann (submitted).

The ellipsoidal correction for the modified Stokes formula is derived in Sjöberg (2003b) and it can be expressed (cf. ibid., eq. 39) as

$$\delta N_{ell}^{Mod} = \frac{R}{4\pi\gamma}\iint_{\sigma_0} S^L(\psi)\left(k\Delta g + \frac{a}{R}\delta g_e\right)d\sigma \ , \qquad (13)$$

where δg_e is the ellipsoidal correction of the gravity anomaly, and is calculated as (ibid., eq. 29e)

$$\delta g_e = e^2\frac{GM}{2a^2}\sum_{n=2}^{\infty}\sum_{m=-n}^{n} Y_{nm}\left\{\left[3-(n+2)F_{nm}\right]C_{nm} - \right. \qquad (14)$$
$$\left. -(n+1)G_{nm}C_{n-2,m} - (n+7)E_{nm}C_{n+2,m}\right\}$$

Note that similarly to Eq. (9), the factor k is also presented in Eq. (13). Exactly the same parameter set s_n used in the geoid determination has also to be utilized in Eq. (13). If any other modification method is applied for Eq. (11), then the respective ellipsoidal correction needs to be derived.

4 Numerical Investigations

In this section the results of the numerical studies of Eqs. (4), (6), (9) and (13), are presented.

For the computation of the ellipsoidal corrections either the gravity anomalies or approximate geoidal heights are required. In this study the spherical harmonic coefficients of the EGM96 (Lemoine et al. 1998) are utilized. Recall that Δg may be related to different surfaces. For instance, the anomaly in Martinec-Grafarend's method is referred to the reference ellipsoid, whereas the anomalies in Eq. (13) are referred to MSL. Although the EGM coefficients are valid on the bounding sphere (with radius a), these are very often utilized for computing Δg and N^0 inside this sphere. For instance, the Δg can be computed by the well-known relationship

$$\Delta g = \frac{GM}{r_e^2}\sum_{n=2}^{360}\left(\frac{a}{r_e}\right)^n (n-1)\sum_{m=-n}^{n} C_{nm}Y_{nm} \ . \qquad (15)$$

Heck and Seitz (2003) point out that Eq. (15) cannot be applied directly, since it might diverge between the bounding sphere and the Earth's ellipsoidal surface. Instead, they suggest the application of some complexly corrected coefficients of T (see ibid., eqs. 18-32) for such a procedure. In this study, however, we thus simply assume that EGM96 is a 'correct' representation of the gravity field (to degree 360 which corresponds to spatial resolution of 55 km) all way down to the Earth's surface. This simplification is sufficient for the practical computations in the global scale, supported also by comparing the results in Table 1 and that of Huang et al. (2003, table 1). The numerical inconsistencies are insignificant and thus do not affect the main scope of the present study. Furthermore, the relatively low resolution of the EGM96 permits us to ignore some minor corrections for gravity anomaly due to separation of the MES and MSL. The above simplifications, of course, may not be allowable when the future high-resolution global data will be employed. A to-date shortage of higher resolution global data certainly overrides the possible contribution of the above approximations. Accordingly, approximate geoidal heights N^0 in Eqs. (6) and (7) are referred to the MES (R = 6371 km) and computed (cf. Fei and Sideris 2000, eq. 52)

$$N^0 = R\sum_{n=2}^{360}\sum_{m=-n}^{n} C_{nm}Y_{nm} \ . \qquad (16)$$

The magnitude of geoidal heights N^0, anomaly Δg, and the integral argument of Eq. (13), range globally from -106 to + 85 m, from -303 to +281 mGal, and between ± 0.3 mGal, respectively. These values on a 30′ x 30′ global grid were exploited for the integration in Eqs. (4), (6) and (13). The coefficients E_{nm}, F_{nm}, G_{nm} for Eqs. (9) and (14) were determined from Sjöberg (2003a, eqs. A6-8).

Table 1 presents the statistics of the reviewed ellipsoidal corrections. Interestingly, a strong correlation between the Martinec-Grafarend ellipsoidal correction and the EGM96 geoid can be observed, see Figure 1. The local extremes of the geoid and ellipsoidal corrections are with the same sign in the same regions, thus the corrections practically mirror the geoid model. Roughly, the magnitude of δN_{ell}^{MG} comprises 0.5-1% of N^0.

A remarkable finding is that the range and global distribution of the Fei-Sideris (2000) and Sjöberg (2003a) ellipsoidal corrections are very similar, see Table 1 and Figure 2. Although their derivational approaches and resulting expressions {cf. Eqs. (6) and (9)} are considerably different, the numerical discrepancies are only about ± 1 cm, globally. This suggests that the two methods practically coincide. Generally, in the area where the gravity anomalies are above MES (i.e. between latitudes ± 45°) the sign of ellipsoidal correction coincides with the one of N^0. In regions where the gravity anomalies are below MES (i.e. in sub-polar latitudes) the ellipsoidal corrections generally have opposite sign to the respective N^0. Intuitively, the absolute values

Table 1. Global range and statistics of different ellipsoidal corrections. Units metre.

Ellipsoidal correction method	max : min : mean : std
Martinec and Grafarend (1998)	+0.83 : -1.09 : -0.00 : 0.30
Fei and Sideris (2000)	+0.47 : -0.62 : +0.01 : 0.16
Sjöberg (2003a)	+0.47 : -0.63 : +0.01 : 0.16
Sjöberg (2003b), modif. Stokes'	+0.01 : -0.01 : +0.00 : 0.00

of geoidal heights increase when the right side of Eq. (1) need to be continued downwards to MES, and diminish when Eq. (1) is continued upwards.

A detailed comparison between the approaches by Molodensky et al. (1962), Moritz (1980), Martinec and Grafarend (1997) and Fei and Sideris (2000) has been published by Huang et al. (2003). The EGM96 was adopted in generating the synthetic data. The outcomes of our study (see Table 1, and Figures 1-2) agree well with the results in Huang et al. (2003, see table 1 and figures 2c,d). They concluded that Martinec-Grafarend's solution demonstrates the best conformity to the synthetic geoid, while the other three methods reveal deficiencies in the first-degree spherical harmonics.

Note that the contribution from the zero- and first-degree effect $\delta N_{0,1}$ by Eq. (9), ranges to ± 5.6 cm, globally. Interestingly, the pattern and range of $\delta N_{0,1}$ in the Sjöberg and Heck-Seitz (2003, fig. 5) methods are essentially the same. The absolute magnitude of the global extremes of the latter $\delta N_{0,1}$ are about 1-2 cm greater than that of Sjöberg's method. Interestingly, regarding to Huang et al. (2003, see fig. 3d), the differences between the synthetic geoid and the geoid computed from the synthetic gravity anomaly for the Fei-Sideris method are rather similar to $\delta N_{0,1}$ pattern (and range) of the Sjöberg and Heck-Seitz approaches.

The range of the ellipsoidal correction for the modified Stokes' formula, see Figure 3, is almost two orders less than the correction for the original Stokes formula. The corrections are largest in equator and in polar regions, as is expected. As the integration in Eq. (13) is carried out within the integration cap (in our computations $\psi_0 = 3°$), in regions with large gravity anomalies the short wavelength features may become relatively significant, though rarely exceeding ± 1 cm, globally. Various sets of least squares parameters $\{(s_n+Q_n^L) \approx 2/(n-1)\}$ were tested. Certainly, variations in any of the input coefficients can slightly change the range of the respective ellipsoidal correction. However, the magnitude of the resulting corrections remains in cm level. It is important to note that higher than 360 degree contribution of global data (Δg and N^0) was not available in this study. Apparently, the

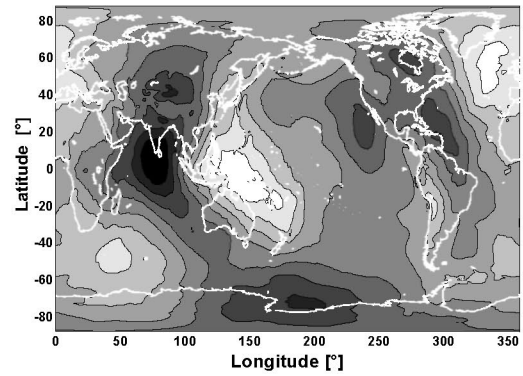


Fig. 1 The global distribution of Martinec-Grafarend's (1997) ellipsoidal correction. The corrections range from +0.83 m (corresponding to the brightest region) to –1.09 m (darkest region), contour interval is 20 cm. Note that the pattern of the corrections practically mirror the EGM96 geoidal heights.

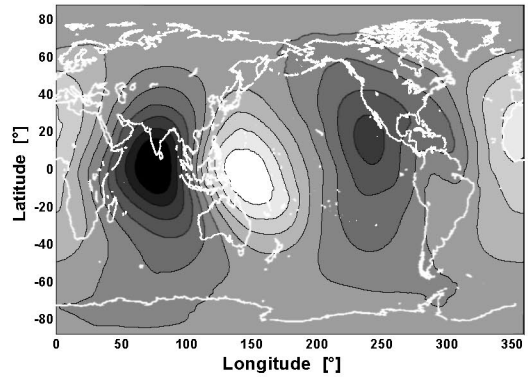


Fig. 2 The global distribution of the Sjöberg (2003a) ellipsoidal correction. The corrections range from + 47 cm (brightest region) to –63 cm (darkest region), contour interval 10 cm. Note the correlation to the geoidal heights in equatorial zone, whereas in sub-polar latitudes the opposite to N^0 sign of the ellipsoidal correction prevails. The Fei-Sideris (2000) ellipsoidal correction practically coincides (within 1 cm) with the Sjöberg correction.

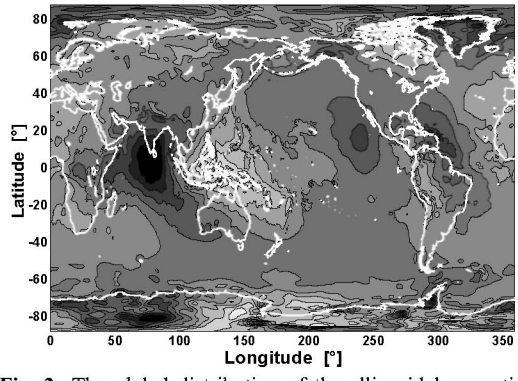


Fig. 3 The global distribution of the ellipsoidal correction for the modified Stokes' formula. The correction ranges between ± 1 cm, globally, contour interval is 2 mm.

ellipsoidal correction for large (and rough) gravity anomaly regions might become notable and, if so, cannot be ignored in precise geoid modelling.

5 Discussion and Summary

Three different methods of the ellipsoidal correction for the original Stokes formula were reviewed and numerically compared in this study. The methods provide the correction as an integral over the full solid angle with various kernel functions. Δg and N^0 are integral arguments in the Martinec-Grafarend and Fei-Sideris methods, respectively, whereas Sjöberg's approach exploits both types of data.

Spherical and ellipsoidal approximations of the geoid are adopted in the derivation of the ellipsoidal corrections. As a consequence, the Martinec-Grafarend (1997) method provides almost twice as large correction values, compared to the other methods {see also results in Huang et al. (2003) and, Heck and Seitz (2003)}. Interestingly, although the expressions of the Sjöberg and Fei-Sideris ellipsoidal corrections are very different, the numerical deviations are only minor (presumably, within the computational noise). At same time, if the term kN^0 in Eq. (9) is excluded, then the ellipsoidal corrections in the Sjöberg and Heck-Seitz methods for degrees $n \geq 2$ are equivalent. Although the zero- and first-degree contribution in the two last methods is accounted differently, the numerical discrepancies reach only about 1-2 cm, globally. In other words, the Sjöberg and Fei-Sideris numerical outcomes are almost the same, whereas the differences between the Sjöberg and Heck-Seitz methods are basically caused by the term kN^0 in Eq. (9).

The reasons for the relatively large Martinec-Grafarend correction values remain unclear, and thus call for further studies. Undoubtedly, the inclusion of continent-wide and reliable sets of GPS-levelling points would be of great use. Apparently, the utilization of high-resolution global data is also advantageous. Present lack of those data sets prevented us to give a outright preference to any of the reviewed methods. However, if the spherical approximation (referred to the MES) in geoid computation is applied, then a clear computational advantage is the spherical harmonic approach of Sjöberg (2003a), since global integration is not required and the singularity problem is eliminated (the same is valid for the Moritz and Heck-Seitz methods too).

The absolute range of the ellipsoidal correction in the least squares modification of Stokes' formula does not exceed 1 cm, globally. This seems to prove the common assumption that the ellipsoidal correction could be neglected in the modified Stokes' formula, if cm accuracy is considered as sufficient. However, in regions with large gravity anomalies this effect should be carefully assessed with high-resolution gravity data.

Acknowledgements. The author thanks Prof. L. Sjöberg for valuable comments to an earlier version of the manuscript. An anonymous reviewer is sincerely thanked for the valuable remarks. This study was supported by the Kami Research Foundation, and Lindstrands' Foundation awarded the IUGG 2003 travel grant. This support is cordially acknowledged.

References

Cruz JY (1986) *Ellipsoidal corrections to potential coefficients obtained from gravity anomaly data on the ellipsoid.* Rep. 371, Dept. of Geodetic Science and Surveying, The Ohio State University, Columbus.

Ellmann A (submitted) On the numerical solution of parameters of the least squares modification of Stokes formula. *Proceed. IAG General Meeting, Sapporo 2003.*

Fei ZL, Sideris MG (2000) A new method for computing the ellipsoidal correction for Stokes' formula. *Journal of Geodesy*, **74**, 223-231 and 671.

Heck B, Seitz K (2003) Solutions of the linearized geodetic boundary value problem for an ellipsoidal boundary to order e^3. *Journal of Geodesy*, **77**, 182-192.

Heiskanen WA, Moritz H (1967) *Physical geodesy.* Freeman, San Francisco.

Huang J, Véronneau M, Pagiatakis SD (2003) On the ellipsoidal correction to the spherical Stokes solution of the gravimetric geoid. *Journal of Geodesy*, **77**, 171-181.

Lemoine FG, Kenyon SC, Factor JK, Trimmer RG, Pavlis NK, Chinn DS, Cox CM, Klosko SM, Luthcke SB, Torrence MH, Wang YM, Williamson RG, Pavlis EG, Rapp RH, Olson TR (1998) *The development of the joint NASA GSFC and NIMA Geopotential Model EGM96.* NASA/TP-1998-206861.

Martinec Z (1998) *Boundary-value problems for gravimetric determination of a precise geoid.* Lecture Notes in Earth Sciences 73, Springer, Berlin Heidelberg New York.

Martinec Z, Grafarend EW (1997) Solution to the Stokes' boundary value problem on an ellipsoid of revolution. *Studia geophysica et geodaetica*, **41**, 103-129.

Molodensky MS, Eremeev VF, Yurkina MI (1962) *Methods for study of the external gravitational field and figure of the Earth.* Israel program for Scientific translations, Jerusalem.

Moritz H (1980) *Advanced physical geodesy.* Herbert Wichmann, Karlsruhe/Abacus Press, Tunbridge Wells.

Sjöberg LE (1991) Refined least-squares modification of Stokes formula. *Manuscripta Geodaetica*, **16**, 367-375.

Sjöberg LE (2003a) Ellipsoidal corrections to order e^2 of geopotential coefficients and Stokes' formula. *Journal of Geodesy*, **77**,148-154.

Sjöberg LE (2003b) A spherical harmonic representation of the ellipsoidal correction to the modified Stokes formula. Submitted to *Journal of Geodesy.*

Sjöberg LE (2003c), The correction to the modified Stokes formula for an ellipsoidal Earth. In Santos M (Ed.): *Honouring the academic life of Petr Vanicek*, UNB Technical Report No 218, New Brunswick, pp 99-110.

SYMPOSIUM G04

General Theory and Methodology

B. Heck

ON TOTAL LEAST-SQUARES ADJUSTMENT WITH CONSTRAINTS

Burkhard Schaffrin
Department of Civil and Environmental Engineering and Geodetic Science,
The Ohio State University, Columbus, OH 43210-1275 (schaffrin.1@osu.edu)

Yaron A. Felus
Department of Surveying Engineering, Ferris State University, Big-Rapids, MI 49307-2291

Abstract. For calibration purposes, oftentimes various datasets are compared in such a way that observations enter the coefficient matrix of a Linear Model ("errors-in-variables"). In such a case, the Total Least-Squares approach would be appropriate that was pioneered by G. Golub and C. van Loan in the early eighties. In essence, rather than solving the usual normal equations system for the estimated parameters, the smallest singular values of a slightly extended system is set to be zero, and its eigenvector is re-scaled to provide the estimated parameter vector. The authors have recently presented their studies that show the potential of this technique to provide improved variograms for geostatistical Kriging applications.

Sometimes, however, stability or slow convergence problems may occur with the algorithm as designed so far. In order to increase the stability, additional parameters could be introduced to represent the functional model under investigation, but with a number of constraints that keep the original redundancy unchanged. In the end, the same Total Least-Squares Fit is supposed to result after fewer iterations from the newly developed scheme that, for the first time, allows the integration of constraints between the parameters, thus solving a case that was long considered "untreatable" by the original TLS algorithm.

Keywords. Total Least-Squares, errors-in-variables, fixed constraints

1. Introduction

Total Least-Squares (TLS) is a relatively new adjustment method of estimating parameters in linear models that include errors in all variables. Specifically, given an overdetermined set of linear equations y $A\xi$ where y is the $n\times 1$ observation vector, A is a given $n\times m$ data matrix, and ξ is the $m\times 1$ vector of unknown parameters, the Total Least-Squares problem is concerned with estimating ξ provided that the number of observations n is larger than the number of parameters to be estimated and given that both the observation vector y and the data matrix A are affected by random errors and need to be adjusted. This problem is different from the classical Least-Squares problem where only the observation vector y is affected by random errors.

The following model, with full-rank matrix A, is assumed:

$$(A - E_A)\cdot\xi - (y - e) = 0,$$
$$E\{[E_A, e]\} = 0,$$
$$C\{E_A, e\} = 0,$$
$$D\{vec[E_A, e]\} = \Sigma_0 \otimes I_n \ , \qquad (1.1)$$

where e and E_A denote a random error vector, resp. matrix. $\Sigma_0 = \sigma_0^{\ 2}\cdot I_{m+1}$ is a (m+1)×(m+1) matrix with an unknown variance component σ_0^2 and given identity matrix I_{m+1}. The symbol $\otimes$ denotes the "Kronecker - Zehfuss product" of matrices, defined by:

$$M \otimes N := [m_{ij}\cdot N]$$
$$\text{for } \mathrm{M} = [\mathrm{m_{ij}}] \text{ and } N \text{ arbitrary;} \qquad (1.2)$$

the "vec" operator stacks one column of a matrix under the other, moving from left to right.

In contrast to the Least-Squares (LS) method, which is based on the minimization of $e^Te = (y - A\xi)^T\cdot(y - A\xi)$ under the condition $E_A := 0$, the (equally weighted) TLS principle is based on minimizing the objective function:

$$e^Te + (vec\,E_A)^T(vec\,E_A) = \min(\xi). \qquad (1.3)$$

When performing an adjustment, it is sometimes necessary to fix some parameters to specific values, or to introduce other (linear) constraints. Then a

constrained adjustment is used to solve this problem in the least-squares sense. In this contribution, a constrained total least-squares solution will be presented, along with an iteration scheme.

The paper is organized as follows: Section 2 develops the Total Least-Squares (TLS) adjustment using both the Euler-Lagrange approach as well as the minimum eigenvalue method as developed more recently in the field of numerical analysis (*Golub and van Loan, 1980; van Huffel and Vandewalle, 1991*). Section 3 introduces the concept of constraints and applies the Euler-Lagrange approach to the TLS-Estimate (CTLSE). Furthermore, an iteration scheme to compute CLTSE is presented and compared with the constrained LS solution. Section 4 concludes the discussion with further research recommendations.

2. Solving the Total Least-Squares problem

2.1 The Euler-Lagrange approach

In order to solve the TLS problem as presented in (1.3) and minimize the respective objective function in view of the model (1.1), we define the Lagrange target function as follows where $e_A := vec(E_A) \sim (0, \sigma_0^2 I_m \otimes I_n)$:

$$\Phi(e, e_A, \lambda, \xi) = e^T \cdot e + e_A^T \cdot e_A + \\ + 2\lambda^T \cdot [y - e - A \cdot \xi + E_A \cdot \xi] \quad (2.1)$$

where λ denotes the n×1 vector of Lagrange multipliers; note that $E_A \cdot \xi = (\xi^T \otimes I_n) \cdot e_A$. Thus the Euler-Lagrange necessary conditions are:

$$\frac{1}{2} \cdot \frac{\partial \Phi}{\partial e} = \tilde{e} - \hat{\lambda} = 0 \quad (2.2)$$

$$\frac{1}{2} \cdot \frac{\partial \Phi}{\partial e_A} = \tilde{e}_A - (\hat{\xi} \otimes I_n) \cdot \hat{\lambda} = 0 \quad (2.3)$$

$$\frac{1}{2} \cdot \frac{\partial \Phi}{\partial \lambda} = y - A \cdot \hat{\xi} - \tilde{e} + \tilde{E}_A \cdot \hat{\xi} = 0 \quad (2.4)$$

$$\frac{1}{2} \cdot \frac{\partial \Phi}{\partial \xi} = -A^T \hat{\lambda} + \tilde{E}_A^T \cdot \hat{\lambda} = 0 \quad (2.5)$$

This system is simplified into:

$$(A^T A) \cdot \hat{\xi} = A^T y + \\ + \hat{\xi} \cdot \underbrace{(\hat{\lambda}^T \hat{\lambda}) \cdot (1 + \hat{\xi}^T \hat{\xi})}_{\hat{v}}, \quad (2.6)$$

$$\hat{\lambda} = (\tilde{e} - \tilde{E}_A \cdot \hat{\xi})(1 + \hat{\xi}^T \cdot \hat{\xi})^{-1} = .$$

$$= (y - A \cdot \hat{\xi})(1 + \hat{\xi}^T \cdot \hat{\xi})^{-1}. \quad (2.7)$$

Note that

$$\hat{v} = \frac{(y - A \cdot \hat{\xi})^T \cdot (y - A \cdot \hat{\xi})}{(1 + \hat{\xi}^T \cdot \hat{\xi})} = \\ = (\hat{\lambda}^T \hat{\lambda})(1 + \hat{\xi}^T \hat{\xi}) = \tilde{e}^T \tilde{e} + \tilde{e}_A^T \tilde{e}_A \hat{\xi} = \\ = \min(\xi) \quad (2.8)$$

is Rayleigh's quotient for the matrix $\begin{bmatrix} A^T A & , & A^T y \\ y^T A & , & y^T y \end{bmatrix}$ with $[\hat{\xi}^T, -1]^T$ as the vector argument. Rayleigh's quotient defines the minimum eigenvalue of the augmented matrix, based on the corresponding eigenvector (see, e.g., *G. Strang, 1988*).

Using these equations the following algorithm had been developed to solve the TLS problem:

1. Compute the LS solution:
$$\hat{\xi}^1 = (A^T A)^{-1} \cdot A^T y \ \textbf{(for } \hat{v}^0 := 0). \quad (2.9a)$$
2. Insert the solution of step (1) as the initial value for the following iterative process:
$$\hat{\xi}^{i+1} = (A^T A)^{-1} \cdot [A^T y + \\ + \hat{\xi}^i \cdot \frac{(y - A \cdot \hat{\xi}^i)^T \cdot (y - A \cdot \hat{\xi}^i)}{(1 + (\hat{\xi}^i)^T \cdot \hat{\xi}^i)}]. \quad (2.9b)$$
3. End when $\left\| \hat{\xi}^{i+1} - \hat{\xi}^i \right\| < \varepsilon$. Then
$$\hat{\sigma}_0^2 = \hat{v} / (n - m). \quad (2.9c)$$

The algorithm seems to converge to the TLS solution in most cases although its efficiency (convergence rate, convergence radius, etc.) still needs to be further investigated. Nevertheless, the establishment of an equivalence between the Euler-Lagrange approach and the minimum eigenvalue method, as will be described in the next subsection, provides an insight into the process that will support future developments.

2.2 The minimum eigenvalue method

In order to solve the TLS problem, equation (1.1) is rewritten in the following form:

$$\begin{bmatrix} A^T - E_A^T \\ y^T - e^T \end{bmatrix} \cdot [A - E_A \quad , \quad y - e] \cdot \begin{bmatrix} \xi \\ -1 \end{bmatrix} =$$

$$= \left(\begin{bmatrix} A^T \cdot A & , & A^T \cdot y \\ y^T \cdot A & , & y^T \cdot y \end{bmatrix} - \Delta \right) \cdot \begin{bmatrix} \xi \\ -1 \end{bmatrix} =$$

$$= \begin{bmatrix} A^T \cdot A & , & A^T \cdot y \\ y^T \cdot A & , & y^T \cdot y \end{bmatrix} \cdot \begin{bmatrix} \xi \\ -1 \end{bmatrix} -$$

$$- \Delta \cdot \begin{bmatrix} \xi \\ -1 \end{bmatrix} = 0 \qquad (2.10)$$

where the error matrix $\Delta = \Delta(E_A, e)$ is unknown. This matrix will be determined such that the condition (1.3) is met using the smallest error matrix Δ as measured by the Frobenius norm:

$$\|\Delta\|_F = [\sum_{i=1}^{m+1} \sum_{j=1}^{m+1} \Delta_{ij}^2]^{1/2} = [tr(\Delta^T \cdot \Delta)]^{1/2} =$$

$$= \min(E_A, e, \xi). \qquad (2.11)$$

In order to satisfy equation (2.10) the following expression should be singular (after applying a SVD decomposition):

$$\begin{bmatrix} A^T \cdot A & , & A^T \cdot y \\ y^T \cdot A & , & y^T \cdot y \end{bmatrix} - \Delta = .$$

$$= [U \quad , \quad u] \cdot \begin{bmatrix} \Lambda & 0 \\ 0 & \lambda_{m+1} \end{bmatrix} \cdot \begin{bmatrix} U^T \\ u^T \end{bmatrix} - \Delta . \qquad (2.12)$$

Here the Singular Value Decomposition was used with the provision that all the eigenvalues in Λ are larger than the smallest eigenvalue λ_{m+1}, and [U, u] contains the corresponding eigenvectors. The minimum error matrix Δ that fulfills equation (2.12), according to the Eckart-Young-Mirsky theorem, is given by:

$$\hat{\Delta}_{\min} = \lambda_{m+1} \cdot (u \cdot u^T) = \Delta([\tilde{E}_A, \tilde{e}]) . \qquad (2.13)$$

Consequently, after removing the smallest singular value, the following system is obtained:

$$(U \cdot \Lambda \cdot U^T) \cdot \begin{bmatrix} \hat{\xi} \\ -1 \end{bmatrix} = 0 ,$$

$$\text{or:} \quad U^T \begin{bmatrix} \hat{\xi} \\ -1 \end{bmatrix} = 0 = U^T u \cdot \qquad (2.14)$$

From (2.14) the estimation of $\hat{\xi}$ is straight-forward by using the fact that the columns of U form an orthogonal complement to the range space of u. Thus, the solution $\hat{\xi}$ is found by scaling this vector until its last component is -1, namely:

$$\hat{\xi}^T = -\frac{1}{u_{m+1}} [u_1 \ldots u_m] . \qquad (2.15)$$

$\lambda_{m \times 1}$ corresponds to $\hat{v}$ in the Euler-Lagrange approach. Thus

$$\tilde{e}^T \cdot \tilde{e} + vec(\tilde{E}_A)^T \cdot vec(\tilde{E}_A) =$$

$$= y^T \cdot y - (A^T y)^T \cdot \hat{\xi} =$$

$$= \lambda_{m+1} = \hat{v} = \hat{\sigma}_0^2 (n-m)^{-1} . \qquad (2.16)$$

In other words, the combined measure-of-fit for the adjustment is expressed by the smallest eigenvalue of the augmented matrix $[A; y]^T[A;y]$ in the case of Total Least-Squares.

3. Total Least-Squares with linear constraints

The following model is the Errors-In-All-Variables Model with Constraints and a full-rank matrix A:

$$y - e = (A_1 - E_1) \cdot \xi_1 + (A_2 - E_2) \cdot \xi_2 ,$$

$$\text{rk}\,[A_1;\ A_2] = m < n , \qquad (3.1)$$

$$\kappa_0 = K_2 \cdot \xi_2 + \underset{l \times l}{K_1} \cdot \xi_1 , \ K_1 \text{ invertible},$$

$$\text{rk}\,[K_1;\ K_2] = l < m , \qquad (3.2)$$

where m is the number of parameters, n the number of independent observation, and l the number of constraints. Part of the parameters can be extracted from the constraints via:

$$\xi_1 = K_1^{-1} (\kappa_0 - K_2 \cdot \xi_2) . \qquad (3.3)$$

By inserting these parameters back to the equation we obtain:

$$[\underbrace{(A - A_1 K_1^{-1} K_2)}_{A'} - \underbrace{(E_2 - E_1 K_1^{-1} K_2)}_{E'}] \xi_2 =$$

$$= \underbrace{(y - A_1 K_1^{-1} \kappa_0)}_{y'} - \underbrace{(e - E_1 K_1^{-1} \kappa_0)}_{e'} . \qquad (3.4)$$

Equation (3.4) leads to a "weighted" TLS problem with the following stochastic characteristics for e' and $vecE_A'$, as derived from the characteristics of e and $vecE_A$ by error propagation:

$$E\{e'\} = 0 ;$$

$$D\{e'\} = \sigma_0^2 [1 + \kappa_0^T (K_1 K_1^T)^{-1} \kappa_0] \cdot I_n ; \qquad (3.5)$$

$$E\{vecE_A'\} = 0 ;$$

$$D\{vec(E_A')\} = \sigma_0^2 [I_{m-l} +$$

$$+ K_2^T (K_1 K_1^T)^{-1} K_2] \otimes I_n ; \qquad (3.6)$$

$$C\{vec(E_A'), e'\} = \sigma_0^2 [K_2^T (K_1 K_1^T)^{-1} \kappa_0] \otimes I_n \quad (3.7)$$

The treatment of "weighted" TLS problems is still very much under investigation; see, e.g., *I. Markovsky et al. (2002)*. We, therefore, follow the direct Euler-Lagrange approach as before.

Incorporating the weight information in the Euler-Lagrange approach is standard procedure as described by *K.R. Koch (1999)*. For this purpose we define the following target function:

$$\Phi(e, e_A, \lambda, \mu, \xi) := e^T e + e_A^T e_A + \\ + 2\lambda^T (y - A\xi - e + E_A \xi) - 2\mu^T (\kappa_0 - K\xi) \quad (3.8)$$

and form the Euler-Lagrange necessary conditions which can be summarized in the following *nonlinear normal equations:*

$$\begin{bmatrix} A^T A & A^T y & K^T \\ y^T A & y^T y & \kappa_0^T \\ K & \kappa_0 & \hat{\nu} I_l \end{bmatrix} \begin{bmatrix} \hat{\xi} \\ -1 \\ \overline{\mu} \end{bmatrix} = \begin{bmatrix} \hat{\xi} \\ -1 \\ \overline{\mu} \end{bmatrix} \cdot \hat{\nu},$$

$$\hat{\nu} = \hat{\nu}_{\min} \geq 0. \quad (3.9)$$

Apparently, this system can also be interpreted as *(implicit) eigenvalue problem* since the minimum nonnegative eigenvalue $\hat{\nu}$ appears inside the matrix as well.

A slowly (but safely) converging algorithm can now be designed as follows (see *B. Schaffrin, 2003*):

1. Start with $\hat{\nu}^0 := 0$ and $\hat{\mu}^0 := 0$ to find the initial solution

$$\begin{bmatrix} \hat{\xi}^1 \\ \overline{\mu}^1 \end{bmatrix} = \begin{bmatrix} A^T A & K^T \\ K & 0 \end{bmatrix}^{-1} \begin{bmatrix} A^T y \\ \kappa_0 \end{bmatrix}. \quad (3.10a)$$

2. Update the coefficient

$$\hat{\nu}^i := \frac{(y - A\hat{\xi}^i)^T (y - A\hat{\xi}^i)}{[1 + (\hat{\xi}^i)^T \hat{\xi}^i]} \quad (3.10b)$$

and compute the new solution

$$\begin{bmatrix} \hat{\xi}^{i+1} \\ \overline{\mu}^{i+1} \end{bmatrix} = \begin{bmatrix} A^T A & K^T \\ K & 0 \end{bmatrix}^{-1} \begin{bmatrix} A^T y + \hat{\xi}^i \hat{\nu}^i \\ \kappa_0 \end{bmatrix} \quad (3.10c)$$

3. End when $\left\| \hat{\xi}^{i+1} - \hat{\xi}^i \right\| < \varepsilon$ and

$$\left\| \hat{\nu}^i - \left(y^T y - c^T \hat{\xi}^i \right) - e_0^T \overline{\mu}^i \right\| < \delta.$$

Then: $\hat{\sigma}_o^2 = \hat{\nu} / (n - m + l)$. (3.10d)

4. Compute, as first order approximation of the dispersion matrix,

$$D\left\{ \begin{bmatrix} \hat{\xi} \\ \overline{\mu} \end{bmatrix} \right\} = \sigma_0^2 \begin{bmatrix} N - \hat{\nu} I_m & K^T \\ K & 0 \end{bmatrix}^{-1} \cdot \\ \cdot \begin{bmatrix} N & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} N - \hat{\nu} I_m & K^T \\ K & 0 \end{bmatrix}^{-1}. \quad (3.10e)$$

For an example, we have to refer the reader to some forthcoming papers; among them will be one on an improved method for fitting variograms based on CTLSE which can be used as an alternative to the 2-stage TLS approach by *Y. Felus and B. Schaffrin (2003)*.

4. Discussion and Conclusions

The Total-Least Squares approach has been derived using the Euler-Lagrange method, and the equivalence to the minimum eigenvalue approach has been presented. An important link between the combined squared error terms and the smallest eigenvalue of the augmented matrix $[A;y]^T[A, y]$ has been established.

Note that a "weighted" TLS problem arises for CTLSE when doing partial elimination as in (3.4). This is not an easy task; some efforts in this direction are reported in the literature, for example *I. Markovsky et al. (2002)*. For the sake of simplicity, we instead followed the traditional Euler-Lagrange approach again and were able to define a set of *nonlinear normal equations* for CTLSE. These normal equations can also be interpreted as *(implicit) eigenvalue problem.*

Often the "Generalized Least-Squares Method" is suggested for solving the TLS problem. However, as will be reported in a future contribution, care must be taken regarding the proper implementation of the algorithm. In fact, many procedures have been reported in the Surveying Engineering literature that will inadvertently (and incorrectly) lead to the ordinary Least-Squares solution.

At this point, it is not known how *sensitive* the TLSS or the CTLSE will be with respect to changes of A, or how restricted the choice of the homoskedastic dispersion model is. Questions like these will be investigated in the future.

Last, this contribution provides the basic mathematical principles required for calibration purposes. Due to the short format of this contribution, we are unable to further develop the

calibration issue, which requires special treatment to preserve the structure of the data matrix.

5. References.

Felus, Y. and B. Schaffrin (2003). Fitting variogram models using the 2-stage Total Least-Squares estimator, *Computers & Geosciences* (submitted in April 2003).

Golub, H.G. and C. van Loan (1980). An analysis of the Total Least-Squares problem, *SIAM Journal on Numerical Analysis*, 17 (6): 883-893.

Koch, K. R. (1999). *Parameter Estimation and Hypothesis Testing in Linear Models*, 2nd edtion. Springer-Verlag, New York, 333 pp.

Markovsky, I., M.-L Rastello, A. Premoli, A. Kukush and S. van Huffel (2002). The element-wise weighted Total Least-Squares problem. *Report No. 02-48*, Dept. of Electrical Eng., Katholic University Leuven, Belgium.

Schaffrin, B. (2003): A note on Constrained Total Squares Estimation, *Comput. Statistics and Data Analysis* (submitted in Oct 2003).

Strang, G. (1988). *Linear Algebra and its Applications*, 3rd edition, Harcourt Brace Jovanovich, San Diego, 505 pp.

van Huffel, S. and J. Vandewalle (1991). *The Total Least-Squares Problem. Computational Aspects and Analysis*, Society for Industrial and Applied Mathematics, Philadelphia, 300 pp.

GNSS Best Integer Equivariant Estimation

P.J.G. Teunissen
Department of Mathematical Geodesy and Positioning
Delft University of Technology
Thijsseweg 11
2629 JA Delft, The Netherlands
Fax: ++ 31 15 278 3711

Abstract The present study is motivated by the problem of GNSS carrier phase ambiguity resolution. We know, when the ambiguity success rate of the usual GNSS model is large enough, that the 'fixed' baseline estimator outperforms its 'float' counterpart in the sense that its probability of being close to the true but unknown baseline is larger than that of the 'float' baseline, provided that the ambiguity success rate is sufficiently large. This is a strong result, but unfortunately only valid if the probability of correct integer estimation is sufficiently close to one. We therefore pose the question whether it is possible to devise a baseline estimator which *always* outperforms its 'float' counterpart. We will show that this is indeed possible, be it that the result comes at the price of having to use a weaker performance criterion. As our main result we present a Gauss-Markov-like theorem which introduces a new minimum variance unbiased estimator that is always superior to the well-known best linear unbiased estimator (BLU) of the Gauss-Markov theorem. This minimum variance unbiased estimator is referred to as the best integer equivariant (BIE) estimator.

Keywords. GNSS ambiguity resolution, best integer equivariant estimation

1 Introduction

Global Navigation Satellite System (GNSS) ambiguity resolution is the process of resolving the unknown cycle ambiguities of double difference (DD) carrier phase data. Its practical importance becomes clear when one realizes the great variety of current and future GNSS models to which it applies. An overview of GNSS models, together with their applications in surveying, navigation, geodesy and geophysics, can be found in textbooks such as [*Hofmann-Wellenhof et al.*, 2001], [*Leick*, 1995], [*Misra and Enge*, 2001], [*Parkinson and Spilker*, 1996], [*Strang and Borre*, 1997] and [*Teunissen and Kleusberg*, 1998].

This study is motivated by the problem of GNSS carrier phase ambiguity resolution. We know, when the ambiguity success rate of the usual GNSS model is large enough, that the 'fixed' baseline estimator $\check{b}$ outperforms its 'float' counterpart $\hat{b}$ in the sense that for the probabilities

$$P(\check{b} \in E_b) \geq P(\hat{b} \in E_b) \tag{1}$$

for any convex region $E_b \subset R^p$ symmetric with respect to the unknown but true baseline $b \in R^p$. This is a strong result, but unfortunately only valid if the probability of correct integer estimation is sufficiently close to one. We therefore raise the question whether it is possible to devise a baseline estimator which *always* outperforms its 'float' counterpart. Our approach to this problem is to opt for a weaker performance criterion and a larger class of estimators to choose from. As performance criterion we choose the mean squared error (MSE).

Our main result is a Gauss-Markov-like theorem which introduces an estimator that is always superior to the well-known best linear unbiased estimator (BLU) of the Gauss-Markov theorem. The Gauss-Markov theorem states that the minimum variance unbiased estimator within the class of linear estimators is given by the least-squares estimator. Our theorem states that the minimum variance unbiased estimator within the class of integer equivariant esti-

mators is given by the least mean-squared estimator. This estimator, referred to as the best integer equivariant (BIE) estimator, is superior to the BLU-estimator since the class of linear unbiased estimators can be shown to be a subset of the class of integer equivariant estimators. In the same sense it can also be shown to be superior to integer estimators.

When the theory is applied to solve for the baseline vector in the GNSS model, we have as a direct consequence that,

$$\begin{aligned} &\mathrm{MSE}(\hat{b}_{BIE}) \leq \mathrm{MSE}(\hat{b}) \text{ and} \\ &\mathrm{MSE}(\hat{b}_{BIE}) \leq \mathrm{MSE}(\check{b}) \end{aligned} \tag{2}$$

Hence the MSE of the BIE-estimator of the baseline is less or at the most equal to that of both its 'float' and 'fixed' counterparts. Although this is a much weaker performance statement than (1), it has the advantage of being valid all the time. In this contribution we will present the results without the rather lengthy proofs, for which the reader is referred to [*Teunissen*, 2003].

2 Integer Estimation

2.1 The GNSS model

As our point of departure we take the following system of linear observation equations

$$E\{y\} = Aa + Bb \ , \ a \in Z^n \ , \ b \in R^p \tag{3}$$

with $E\{.\}$ the mathematical expectation operator, y the m-vector of observables, a the n-vector of unknown integer parameters and b the p-vector of unknown real-valued parameters. The $m \times (n+p)$ design matrix (A, B) is assumed to be of full rank.

All the linear(ized) GNSS models can in principle be cast in the above frame of observation equations. The data vector y will then usually consist of the 'observed minus computed' single- or dual- frequency double-difference (DD) phase and/or pseudo-range (code) observations accumulated over all observation epochs. The entries of vector a are then the DD carrier phase ambiguities, expressed in units of cycles rather than range, while the entries of the vector b will consist of the remaining unknown parameters, such as for instance baseline components (coordinates) and possibly atmospheric delay parameters (troposphere, ionosphere).

The procedure which is usually followed for solving the GNSS model can be divided into three steps. In the *first* step one simply discards the integer constraints $a \in Z^n$ and performs a standard least-squares (LS) adjustment. As a result one obtains the LS-estimator of a and b together with their variance-covariance (vc-) matrix as

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \ , \ \begin{bmatrix} Q_{\hat{a}} & Q_{\hat{a}\hat{b}} \\ Q_{\hat{b}\hat{a}} & Q_{\hat{b}} \end{bmatrix} \tag{4}$$

This solution is usually referred to as the 'float' solution.

In the *second* step the 'float' estimator $\hat{a}$ is used to compute the corresponding integer estimator $\check{a} \in Z^n$. This implies that a mapping S from the n-dimensional space of reals to the n-dimensional space of integers is introduced such that

$$\check{a} = S(\hat{a}) \ , \ S : R^n \mapsto Z^n \tag{5}$$

This integer estimator is then used in the final and *third* step to adjust the 'float' estimator $\hat{b}$. As a result one obtains the so-called 'fixed' estimator of b as

$$\check{b} = \hat{b} - Q_{\hat{b}\hat{a}} Q_{\hat{a}}^{-1} (\hat{a} - \check{a}) \tag{6}$$

In the following we will use the terminology of GNSS and refer to $\hat{b}$ and $\check{b}$ as respectively the 'float' and 'fixed' baseline estimators.

The above three-step procedure is still ambiguous in the sense that it leaves room for choosing the integer map S. Different choices for S will lead to different integer estimators $\check{a}$ and thus also to different baseline estimators $\check{b}$. One can therefore now think of constructing integer maps which possess certain desirable properties.

2.2 A class of integer estimators

It will be clear that the map S will not be one-to-one due to the discrete nature of Z^n. Instead it will be a many-to-one map. This implies that different real-valued vectors will be mapped to one and the same integer vector. One can therefore assign a subset $S_z \subset R^n$ to each integer vector $z \in Z^n$:

$$S_z = \{x \in R^n \mid z = S(x)\}, \quad z \in Z^n \tag{7}$$

The subset S_z contains all real-valued vectors that will be mapped by S to the same integer vector $z \in Z^n$. This subset is referred to as the *pull-in region* of z. It is the region in which all vectors are pulled to the same integer vector z.

Since the pull-in regions define the integer estimator completely, one can define classes of integer estimators by imposing various conditions on the pull-in regions. One such class was introduced by Teunissen (1999a) and is referred to as the class of admissible integer estimators.

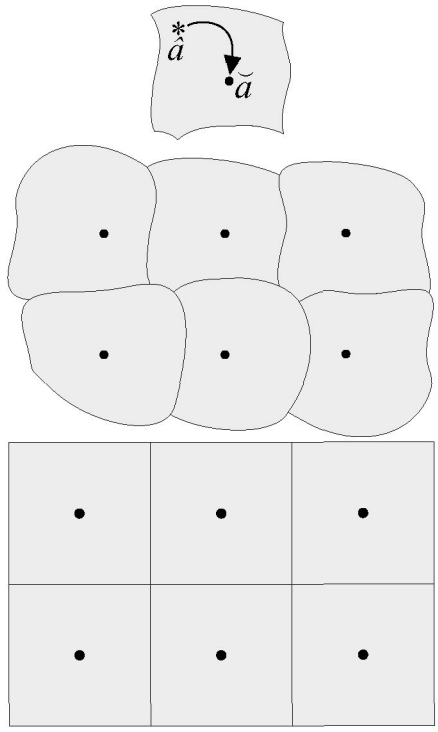


Fig. 1 Pull-in regions. Top: arbitrary pull-in region of $z = \check{a}$; Middle: pull-in regions that cover R^n without gaps and overlaps; Bottom: an example of integer translational invariant pull-in regions that cover R^n without gaps and overlaps.

Definition 1 (*admissible integer estimators*)
The integer estimator $\check{a} = S(\hat{a})$ is said to be *admissible* if its pull-in regions satisfy

$$
\begin{aligned}
&(i) \quad \bigcup_{z \in Z^n} S_z = \mathbb{R}^n \\
&(ii) \quad \mathrm{Int}(S_{z_1}) \bigcap \mathrm{Int}(S_{z_2}) = \emptyset, \quad \forall z_1, z_2 \in \mathbb{Z}^n, \; z_1 \neq z_2 \\
&(iii) \quad S_z = z + S_0, \quad \forall z \in \mathbb{Z}^n
\end{aligned}
$$

This definition is motivated as follows (see Figure 1). Each one of the above three conditions describe a property of which it seems reasonable that it is possessed by an arbitrary integer estimator. The first condition states that the pull-in regions should not leave any gaps and the second that they should not overlap. The absence of gaps is needed in order to be able to map any 'float' solution $\hat{a} \in R^n$ to Z^n, while the absence of overlaps is needed to guarantee that the 'float' solution is mapped to just one integer vector. Note that we allow the pull-in regions to have common boundaries. This is permitted if we assume to have zero probability that $\hat{a}$ lies on one of the boundaries. This will be the case when the probability density function (pdf) of $\hat{a}$ is continuous.

The third and last condition of the definition follows from the requirement that $S(x + z) = S(x) + z, \forall x \in R^n, z \in Z^n$. Also this condition is a reasonable one to ask for. It states that when the 'float' solution $\hat{a}$ is perturbed by $z \in Z^n$, the corresponding integer solution is perturbed by the same amount. This property allows one to apply the *integer remove-restore* technique: $S(\hat{a} - z) + z = S(\hat{a})$. It therefore allows one to work with the fractional parts of the entries of $\hat{a}$, instead of with its complete entries.

Using the pull-in regions, one can give an explicit expression for the corresponding integer estimator $\check{a}$. It reads

$$\check{a} = \sum_{z \in Z^n} z s_z(\hat{a}) \tag{8}$$

with the indicator function $s_z(\hat{a}) = 1$ if $\hat{a} \in S_z$ and $s_z(\hat{a}) = 0$ otherwise. Note that the $s_z(\hat{a})$ can be interpreted as weights, since $\sum_{z \in Z^n} s_z(\hat{a}) = 1$. The integer estimator $\check{a}$ is therefore equal to a weighted sum of integer vectors with binary weights.

2.3 The performance of the baseline estimators

As mentioned earlier the 'fixed' baseline estimator $\check{b}$ depends on the chosen integer estimator $\check{a}$. We may write $\check{b}$ also as

$$\check{b} = \hat{b}(a) + Q_{\hat{b}\hat{a}} Q_{\hat{a}}^{-1} (\check{a} - a) \tag{9}$$

with the conditional baseline estimator $\hat{b}(a) = \hat{b} - Q_{\hat{b}\hat{a}} Q_{\hat{a}}^{-1} (\hat{a} - a)$. Since a is assumed to be known in case of $\hat{b}(a)$, the conditional baseline estimator is the best possible estimator of b. It is unbiased, it has a precision which is better than that of $\hat{b}$, and in case y is normally distributed its pdf is also more peaked than that of the 'float' estimator. The best one can hope for in case of $\check{b}$ is therefore that $\check{a} = a$. But this requires that the probability of correct integer estimation, $P(\check{a} = a)$, equals one. Hence, of all admissible integer estimators the preferred estimator is the one which maximizes the probability of correct integer estimation. It was shown by Teunissen (1999b) that in case of elliptically contoured distributions, the integer least-squares (ILS) estimator is the preferred estimator. Thus, if $\check{a}_{ILS} = \arg\min_{z \in Z^n} (\hat{a} - z)^T Q_{\hat{a}}^{-1} (\hat{a} - z)$, then $P(\check{a}_{ILS} = a) \geq P(\check{a} = a)$ for

any admissible estimator $\check{a}$. And once the probability of correct integer estimation is sufficiently close to one, we have

$$P(\check{b} \in E_b) \geq P(\hat{b} \in E_b) \tag{10}$$

for any convex region $E_b \subset R^p$ symmetric with respect to b, see [*Teunissen*, 2003]. The approach taken with GNSS is therefore usually to use the ILS-estimator for computing the 'fixed' baseline estimator, once it has been verified that $P(\check{a}_{ILS} = a)$ is sufficiently close to one.

When valid, inequality (10) is a very strong result. It states that the 'fixed' baseline estimator $\check{b}$ has a higher probability of being close to b than its 'float' counterpart. The downside of the inequality is however that it is ony valid if $P(\check{a} = a)$ is sufficiently close to one. This being the case, one may wonder whether it would not be possible to devise a baseline estimator which *always* outperforms its 'float' counterpart. Such an approach can however only be successful if we use a weaker performance criterion than that of (10). Furthermore, assuming that we will be successful in finding an optimal estimator using this weaker criterion, then this new estimator will always be better than its 'float' counterpart only if we consider a class of estimators which encompasses the class of estimators in which the 'float' estimator resides. But this means that, as a start, we should at least consider a class of estimators which is larger than the above considered class of integer estimators.

3 Integer Equivariant Estimation

3.1 Class of integer equivariant estimators

We will now introduce a new class of estimators which is larger than the previous defined class of integer estimators. And in order to be general enough, we consider estimating an arbitrary linear function of the two types of unknown parameters of the GNSS model (3),

$$\theta = l_a^T a + l_b^T b \ , \ l_a \in R^n \ , \ l_b \in R^p \tag{11}$$

Thus if $l_b = 0$ then linear functions of the ambiguities are estimated, whereas if $l_a = 0$ then linear functions of the baseline are estimated. Linear functions of both the ambiguities and baseline, such as carrier phases, are estimated in case $l_a \neq 0$ and $l_b \neq 0$.

It seems reasonable that the estimator should at least obey the *integer remove-restore* principle. When estimating ambiguities in case of GNSS for instance, one would like, when adding an arbitrary number of cycles to the carrier phase data, that the solution of the integer ambiguities gets shifted by the same integer amount. For the estimator of θ this would mean that adding Az to y, with arbitrary $z \in Z^n$, must result in a shift of $l_a^T z$. Likewise it seems reasonable to require of the estimator that adding $B\zeta$ to y, with arbitrary $\zeta \in R^p$, results in a shift $l_b^T \zeta$. Afterall we would not like the integer part of the estimator to become contaminated by such an addition to y. Estimators of θ that fulfil these two conditions will be called *integer equivariant* (IE). Hence, they are defined as follows

Definition 2 (*integer equivariant estimators*)
The estimator $\hat{\theta}_{IE} = f_\theta(y)$, with $f_\theta : R^m \mapsto R$, is said to be an *integer equivariant* estimator of $\theta = l_a^T a + l_b^T b$ if

$$\begin{cases} f_\theta(y + Az) = f_\theta(y) + l_a^T z, & \forall y \in R^m, z \in Z^n \\ f_\theta(y + B\zeta) = f_\theta(y) + l_b^T \zeta, & \forall y \in R^m, \zeta \in R^p \end{cases} \tag{12}$$

It is not difficult to verify that the integer estimators of the previous section are integer equivariant. Simply check that the above two conditions are indeed fulfilled by the estimator $\check{\theta} = l_a^T \check{a} + l_b^T \check{b}$. The converse, however, is not necessarily true. The class of IE-estimators is therefore indeed a larger class.

One can also show that the class of IE-estimators is larger than the class of linear unbiased estimators. That is, the condition of linear unbiasedness is more restrictive than the condition of integer equivariance. Hence, the class of linear unbiased estimators is a subset of the class of integer equivariant estimators. This implies that IE-estimators exist which are unbiased. Thus, if we denote the class of IE-estimators as IE, the class of unbiased estimators as U, the class of unbiased IE-estimators as IEU, the class of unbiased integer estimators as IU, and the class of linear unbiased estimators as LU, we may summarize their relationships as: $IEU = IE \cap U \neq \emptyset$, $LU \subset IEU$ and $IU \subset IEU$.

4 Best Integer Equivariant Estimation

4.1 The BIE-estimator

Having defined the class of IE-estimators we will now look for an IE-estimator which is 'best' in a certain sense. We will denote our *best integer equivariant* (BIE) estimator of θ as $\hat{\theta}_{BIE}$ and use the mean

squared error (MSE) as our criterion of 'best'. The best integer equivariant estimator will therefore be defined as

$$\hat{\theta}_{BIE} = \arg \min_{f_\theta \in IE} E\{(f_\theta(y) - \theta)^2\} \tag{13}$$

in which IE stands for the class of IE-estimators. The minimization is thus taken over all integer equivariant functions that satisfy the conditions of Definition 2.

The reason for choosing the MSE-criterion is twofold. First, it is a well-known probabilistic criterion for measuring the closeness of an estimator to its target value, in our case θ. Second, the MSE-criterion is also often used as a measure for the quality of the 'float' solution itself. The following theorem gives the solution to the above minimization problem (13).

Theorem 1 (*best integer equivariant estimation*) Let $y \in R^m$ have mean $E\{y\} = Aa + Bb$ and pdf $p_y(y)$, and let $\hat{\theta}_{BIE}$ be the best integer equivariant estimator of $\theta = l_a^T a + l_b^T b$. Then

$$\hat{\theta}_{BIE} = \frac{\sum_{z \in Z^n} \int_{R^p} (l_a^T z + l_b^T \beta) p_y(Y) d\beta}{\sum_{z \in Z^n} \int_{R^p} p_y(Y) d\beta} \tag{14}$$

with $Y = y + A(a - z) + B(b - \beta)$.

Note that the BIE-estimator can also be written as

$$\hat{\theta}_{BIE} = l_a^T \hat{a}_{BIE} + l_b^T \hat{b}_{BIE} \tag{15}$$

with

$$\begin{cases} \hat{a}_{BIE} = \sum_{z \in Z^n} z w_z(y), & \sum_{z \in Z^n} w_z(y) = 1 \\ \hat{b}_{BIE} = \int_{R^p} \beta w_\beta(y) d\beta, & \int_{R^p} w_\beta(y) d\beta = 1 \end{cases}$$

in which the weighting functions $w_z(y)$ and $w_\beta(y)$ are defined by (14). This shows that the BIE-estimator of the integer parameter vector a is also a weighted sum of all integer vectors in Z^n, just like $\check{a}$ of (8) is. In the present case, however, the weights are not binary. They vary between zero and one, and their values are determined by y and its pdf. As a consequence the estimator $\hat{a}_{BIE}$ will be real-valued in general, instead of integer-valued.

The above theorem holds true for any pdf the vector of observables y might have. This is therefore a very general result indeed. A closer look at (14) reveals however, that one needs a and b, and therefore θ, in order to compute $\hat{\theta}_{BIE}$. The dependence on a and b is present in the numerator of (14) and not in its denominator. The summation over all integer vectors in Z^n and the integration over R^p makes the dependence on a and b disappear in the denominator. If the dependence of $\hat{\theta}_{BIE}$ on θ persists one would not be able to compute the BIE-estimator. Note however that this dependence disappears in case the pdf of y has the structure $p_y(y) = f(y - Aa - Bb)$. And this property is fortunately still true for a large class of probability density functions.

4.2 BIE is better than BLU

A direct and important consequence of the above theorem is that the BIE-estimator is always better than or at least as good as any integer estimator as well as any linear unbiased estimator. Afterall the class of integer estimators and the class of linear unbiased estimators are both subsets of the class of IE-estimators. The BIE-estimator is therefore also better in the MSE-sense than the best linear unbiased (BLU) estimator. Moreover it can be shown that the BIE-estimator is unbiased. We therefore have the following Gauss-Markov-like theorem.

Theorem 2 (*minimum variance unbiased estimation*) The BIE-estimator is unbiased and has a better precision than the BLU-estimator:

$$\begin{aligned} &(i) \quad E\{\hat{\theta}_{BIE}\} = E\{\hat{\theta}_{BLU}\} \\ &(ii) \quad D\{\hat{\theta}_{BIE}\} \leq D\{\hat{\theta}_{BLU}\} \end{aligned} \tag{16}$$

where $D\{.\}$ denotes the dispersion operator.

The above result is remarkable since it shows that for a large class of pdf's of y, one can *always*, with a model like (3), improve upon the precision of the BLU-estimator while keeping the estimator unbiased. If we apply the above theorem to the problem of estimating the baseline in case of GNSS and make the comparison with the 'float' baseline estimator and an unbiased 'fixed' baseline estimator, we have

$$\begin{cases} D\{\hat{b}_{BIE}\} \leq D\{\check{b}\} & \text{and} \quad E\{\hat{b}_{BIE}\} = E\{\check{b}\} \\ D\{\hat{b}_{BIE}\} \leq D\{\hat{b}\} & \text{and} \quad E\{\hat{b}_{BIE}\} = E\{\hat{b}\} \end{cases} \tag{17}$$

The above matrix inequalities imply that $l^T D\{\hat{b}_{BIE}\} l \leq l^T D\{\check{b}\} l$ and $l^T D\{\hat{b}_{BIE}\} l \leq l^T D\{\hat{b}_{BLU}\} l$ hold true for all $l \in R^p$. The precision of the baseline estimator $\hat{b}_{BIE}$ is therefore always better than or at least as good as the precision of its 'float' and 'fixed' counterparts.

We are now also in a position to make a comparison with our earlier result (10). It will be clear that the performance statement of (17) is weaker than that of (10). On the other hand, (10) is only valid in case the probability of correct integer estimation is sufficiently close to one, whereas the above result is always valid.

4.3 The Gaussian case

So far we did not make a particular choice for the pdf of y. In many applications however, such as GNSS, it is often assumed that y is normally distributed. In that case the pdf of y takes the form

$$p_y(y) = \frac{1}{(2\pi)^{\frac{m}{2}}\sqrt{\det Q_y}} \times \exp\left\{-\frac{1}{2} \parallel y - Aa - Bb \parallel^2_{Q_y}\right\} \tag{18}$$

where $\| \cdot \|^2_{Q_y} = (.)^T Q_y^{-1}(.)$. With this Gaussian pdf the BIE-estimator also takes on a particular shape. We have the following corollary.

Corollary (*BIE in the Gaussian case*)
Let the pdf of y be given as in (18) and let $\hat{\theta}_{BIE}$ be the best integer equivariant estimator of $\theta = l_a^T a + l_b^T b$. Then

$$\hat{\theta}_{BIE} = l_a^T \hat{a}_{BIE} + l_b^T \hat{b}_{BIE} \tag{19}$$

with

$$\begin{cases} \hat{a}_{BIE} &= \sum\limits_{z \in Z^n} z \dfrac{\exp -\frac{1}{2}\|\hat{a}-z\|^2_{Q_{\hat{a}}}}{\sum\limits_{z \in Z^n} \exp -\frac{1}{2}\|\hat{a}-z\|^2_{Q_{\hat{a}}}} \\ \hat{b}_{BIE} &= \hat{b} - Q_{\hat{b}\hat{a}} Q_{\hat{a}}^{-1}(\hat{a} - \hat{a}_{BIE}) \end{cases}$$

This result shows that in the Gaussian case we may use the three-step procedure of Section 2 also for best integer equivariant estimation. The only difference is that we need to replace the integer estimator $\check{a}$ by the BIE-estimator $\hat{a}_{BIE}$.

It is interesting to observe that the above expressions given for $\hat{a}_{BIE}$ and $\hat{b}_{BIE}$ are identical to their Bayesian counterparts as given in [*Betti et al.*, 1993] and [*Gundlich and Koch*, 2001], see also [*Teunissen*, 2001]. This is not quite true for the general case however. Still, the above equivalence nicely bridges the gap which existed so far between the current theory of integer inference and the Bayesian approach. Despite the similarity in the above case however, there are important differences in the probabilistic evaluation of the solutions. Like the BLU-estimator, the BIE-estimator is a random variable with the property of being unbiased and of minimum variance. In the Bayesian framework the solution is considered to be nonrandom due to the conditioning that takes place. Furthermore, in the Bayesian framework the unknown parameters are assumed to be random variables for which probability distributions need to be specified a priori. The approach in the present contribution is nonBayesian throughout with no need at all to make assumptions about prior distributions.

5 References

Betti B, Crespi M, Sanso F (1993): A geometric illustration of ambiguity resolution in GPS theory and a Bayesian approach. *Manuscr Geod*, 18: 317-330.

Gundlich B, Koch K-R (2002): Confidence regions for GPS baselines by Bayesian statistics. *J Geod*, in print.

Hofmann-Wellenhof B, Lichtenegger H, Collins J (2001): *Global Positioning System: Theory and Practice*. 5th edition. Springer Verlag.

Leick A (1995): *GPS Satellite Surveying*. 2nd edition, John Wiley, New York.

Misra P, Enge P (2001): *Global Positioning System: Signals, Measurements, and Performance*. Ganga-Jamuna Press, 2001.

Parkinson B, Spilker JJ (eds) (1996): *GPS: Theory and Applications*, Vols 1 and 2, AIAA, Washington DC.

Strang G, Borre K (1997): *Linear Algebra, Geodesy, and GPS*, Wellesley-Cambridge Press.

Teunissen PJG, Kleusberg A (eds) (1998): *GPS for Geodesy*, 2nd enlarged edition, Springer Verlag.

Teunissen PJG (1999a): The probability distribution of the GPS baseline for a class of integer ambiguity estimators. *Journal of Geodesy*, 73: 275-284.

Teunissen PJG (1999b): An optimality property of the integer least-squares estimator. *Journal of Geodesy*, 73: 587-593.

Teunissen PJG (2001): Statistical GNSS carrier phase ambiguity resolution: a review. Invited paper in *Proc. of IEEE Symposium Statistical Signal Processing*, Singapore, pp. 4-12.

Teunissen PJG (2003): Theory of integer equivariant estimation with application to GNSS. *Journal of Geodesy*, 77: 402-410.

Performance comparison of the BIE estimator with the float and fixed GNSS ambiguity estimators

S. Verhagen and P.J.G. Teunissen

Department of Mathematical Geodesy and Positioning
Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands
E-mail: S.Verhagen@LR.TUDelft.nl

Abstract. The goal of ambiguity resolution is to make optimal use of the integerness of the ambiguities, and it is the key to high precision GNSS positioning and navigation. However, it should only be applied in case the probability of correct integer ambiguity resolution, i.e. the success rate, is very close to one. In that case, the probability that the fixed baseline will be closer to the true but unknown baseline is larger than that of the float baseline. Clearly, this condition will not be fulfilled for each measurement scenario, and this means that for low success rates a user will prefer the float solution.

However, there exists a baseline estimator that will always be superior to its float and fixed counterparts, albeit that this superiority is measured using a weaker optimality condition. This baseline estimator is the Best Integer Equivariant (BIE) estimator, which is unbiased and of minimum variance within the class of integer equivariant estimators.

In this contribution, the three different estimators are compared. For that purpose, we will focus on the geometry-free GNSS models, either single frequency or dual frequency. The performance of the estimators is compared based on their probability density functions, the variances of the different estimators, and the probabilities that the baseline estimators are within a certain convex region symmetric with respect to the true baseline. This will provide information on whether or not the BIE estimator could be useful in positioning applications.

Keywords. GNSS ambiguity resolution, best integer equivariant unbiased estimator, float and fixed estimator

1 Introduction

A GNSS model generally contains real-valued and integer-valued parameters. The latter are the double difference (DD) carrier phase ambiguities, a. The first group is referred to as the baseline unknowns, b.

The 'float' estimators of the unknown parameters and their variance-covariance (vc-) matrix are obtained after a standard least-squares adjustment, ignoring the integer-constraint on the ambiguity unknowns:

$$\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix};\quad \begin{pmatrix}Q_{\hat{a}} & Q_{\hat{a}\hat{b}}\\ Q_{\hat{a}\hat{b}} & Q_{\hat{b}}\end{pmatrix} \tag{1}$$

In the next step, the float ambiguities are fixed to integer values, which is referred to as ambiguity resolution:

$$\check{a} = S(\hat{a}) \tag{2}$$

where $S : \mathbb{R}^n \mapsto \mathbb{Z}^n$ is the mapping from the n-dimensional space of real numbers to the n-dimensional space of integers. The optimal result in the sense of maximizing the probability of correct integer estimation (success rate) is obtained using integer least-squares, cf (Teunissen, 1993; Teunissen, 1999). Finally, the float baseline estimator is adjusted by virtue of its correlation with the ambiguities. The result is the so-called 'fixed' baseline estimator:

$$\check{b} = \hat{b} - Q_{\hat{b}\hat{a}}Q_{\hat{a}}^{-1}(\hat{a} - \check{a}) \tag{3}$$

Ambiguity resolution should only be applied when there is enough confidence in its results, which means that the success rate should be very close to one. Only then the probability that the fixed baseline is close to the true baseline is higher than the corresponding probability for the float baseline. If this is

not the case, a user will prefer the float solution. So, the choice between float or fixed baseline solution depends on the success rate. However, there exists a baseline estimator that will always be superior to its float and fixed counterparts, albeit that this superiority is measured using a weaker optimality criterion. This baseline estimator is the Best Integer Equivariant (BIE) estimator and it was introduced in (Teunissen, 2003b; Teunissen, 2003a).

2 The BIE estimator

The best integer equivariant (BIE) estimator is based on a new class of estimators. Estimators in this new class only have to fulfill the integer remove-restore property, which requires that $S(\hat{a}-z)+z=S(\hat{a})$. Hence the name integer equivariant (IE) that is assigned to all estimators that belong to this class. The criterion of 'best' that will be used is that the mean squared error (MSE) should be minimal within the IE-class. With this MSE-criterion the best integer equivariant (BIE) estimator is defined as:

$$\hat{\theta}_{\text{BIE}} = \arg \min_{f_\theta \in IE} E\{(f_\theta(y)-\theta)^2\} \tag{4}$$

The minimization is taken over all IE functions $f_\theta(y)$ with $\theta = l_a^T a + l_b^T b$, and $l_a \in \mathbb{R}^n$, $l_b \in \mathbb{R}^p$ are chosen linear functions.

In our GNSS applications we assume that the data are normally distributed. It can be shown that in this case the BIE estimator $\hat{\theta}_{BIE}$ becomes

$$\hat{\theta}_{\text{BIE}} = l_a^T \bar{a} + l_b^T \bar{b} \tag{5}$$

with

$$\begin{cases} \bar{a} = \sum\limits_{z\in\mathbb{Z}^n} z \dfrac{\exp\{-\frac{1}{2}\|\hat{a}-z\|^2_{Q_{\hat{a}}}\}}{\sum\limits_{z\in\mathbb{Z}^n}\exp\{-\frac{1}{2}\|\hat{a}-z\|^2_{Q_{\hat{a}}}\}} \\ \bar{b} = \hat{b} - Q_{\hat{b}\hat{a}} Q_{\hat{a}}^{-1}(\hat{a}-\bar{a}) \end{cases} \tag{6}$$

This shows that $\bar{a}$ is a weighted sum of all integer vectors and could thus also be written as

$$\bar{a} = \sum_{z\in\mathbb{Z}^n} z w_z(\hat{a}) \tag{7}$$

with $w_z(\hat{a}) \leq 1, \forall z \in \mathbb{Z}^n$ and $\sum\limits_{z\in\mathbb{Z}^n} w_z(\hat{a}) = 1$.

Note that the formal expressions in (6) are identical to their Bayesian counterparts as given in (Betti et al., 1993; Gundlich and Koch, 2002; Gundlich, 2002), but that the distributional properties of the BIE estimator and its Bayesian counterpart of course differ. Since the functional form of the nonBayesian estimator is identical to the Bayesian solution in the special case of normally distributed data, the theory of BIE-estimation has provided the as yet unknown link with the Bayesian approach of ambiguity resolution.

In (Teunissen, 2003b) it was also shown that

$$\begin{cases} D\{\bar{b}\} \leq D\{\check{b}\} \text{ and } E\{\bar{b}\} = E\{\check{b}\} \\ D\{\bar{b}\} \leq D\{\hat{b}\} \text{ and } E\{\bar{b}\} = E\{\hat{b}\} \end{cases} \tag{8}$$

where $E\{\cdot\}$ is the mathematical expectation operator and $D\{\cdot\}$ the dispersion operator. Eq.(8) shows that the BIE estimator is always 'better' than the float and fixed counterparts in terms of the variance.

3 Comparison of the estimators

Although we know that the BIE estimator outperforms its float and fixed counterparts in terms of precision, it is not clear how large this difference will be under varying measurement scenarios. It is therefore also of interest to compare the three estimators numerically. This also provides the possibility to compare their distributional properties.

Fixed versus BIE ambiguity estimator

Both the fixed and the BIE ambiguity estimator are weighted sums of all integer vectors in $\mathbb{Z}^n$. However, in case of the fixed integer least-squares estimator a weight of 1 is assigned to the integer vector with minimal distance to the float vector, and all other weights are set to zero. In case of the BIE estimator the weights depend on the distance, again measured in the metric of $Q_{\hat{a}}$, of the integer vectors to the float vectors, see Eq.(6). In general the BIE ambiguity estimates will therefore be real-valued.

It follows that in the limits of the precision the following is true:

$$\lim_{\sigma\to\infty} \bar{a} = \hat{a} \quad \text{and} \quad \lim_{\sigma\to 0} \bar{a} = \check{a} \tag{9}$$

with the vc-matrix factored as $Q_{\hat{a}} = \sigma^2 G$. This shows that if the precision is high and thus the success rate is close to one, the BIE estimator will be close to the fixed estimator. On the other hand, if the precision is low the BIE estimator will approximate the float estimator. It is thus especially interesting to know how the BIE estimator performs in the intermediate cases compared to the float and fixed estimators.

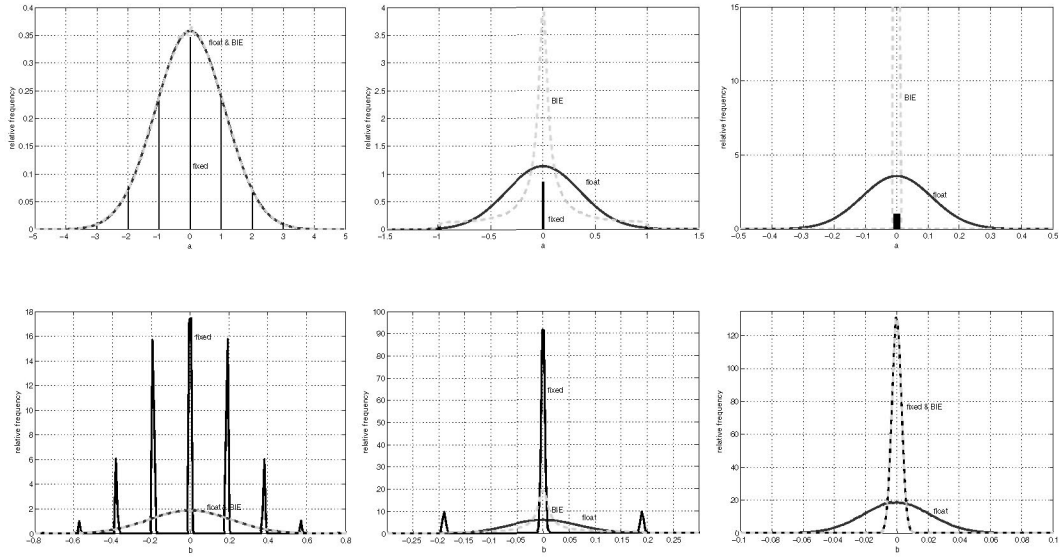


Fig. 1 Examples of distributions of the different estimators. Top: $\hat{a}$ (solid), $\check{a}$ (bars), $\bar{a}$ (dashed); Bottom: $\hat{b}$ (solid), $\check{b}$ (solid, multi-modal), $\bar{b}$ (dashed). Left: 2 epochs; Center: 20 epochs; Right: 200 epochs. Note the different scalings.

Comparison in the one-dimensional case

A random generator was used to generate 500,000 samples of the float range and ambiguity, using the geometry-free single frequency GPS model for k epochs, with vc-matrix:

$$\begin{pmatrix} Q_{\hat{a}} & Q_{\hat{a}\hat{b}_k} \\ Q_{\hat{b}_k\hat{a}} & Q_{\hat{b}_k} \end{pmatrix} = \begin{pmatrix} \frac{\sigma_p^2}{k\lambda^2}(1+\varepsilon) & -\frac{\sigma_p^2}{k\lambda^2} \\ -\frac{\sigma_p^2}{k\lambda^2} & \frac{\sigma_p^2}{1+\varepsilon}(\frac{1}{k}+\varepsilon) \end{pmatrix}$$

with λ the wavelength of the carrier; σ_p^2 and σ_ϕ^2 are the variances of the DD code and phase observations respectively, and $\varepsilon = \frac{\sigma_\phi^2}{\sigma_p^2}$. For all simulations, the standard deviations were chosen as $\sigma_p = 30$ cm and $\sigma_\phi = 3$ mm. The number of epochs was varied. Note that in the one-dimensional case the fixed ambiguity estimator is obtained by simply rounding the float estimator to the nearest integer.

Figure 1 shows the parameter distributions of all three estimators for different values of k, based on the simulation results. It can be seen that for small k the distribution of the BIE ambiguity and range estimator resembles the normal PDF of the float estimators. For larger k, and thus higher precision, the distribution of the BIE estimators more and more resemble those of the fixed estimators. Note that the multi-modality of the distribution of the BIE range estimator is less pronounced than that of the fixed range estimator.

From Eq.(8) follows that the BIE baseline estimator has smallest variance, but in the limits of the precision the variance will become equal to the variance of the float and fixed estimator. This is illustrated in Figure 2, where the variance ratio of the different estimators is shown as function of k. Also the success rate is shown. Indeed, for small k the variance of the BIE and the float estimator are equal to each other (ratio equals one), and smaller than the variance of the fixed estimator. On the other hand, for large k the variances of the BIE and fixed estimator become equal to each other, and smaller than that of the float estimator. Only after 30 epochs, when the success rate is larger than 0.9, the variance of the fixed baseline estimator becomes lower than the variance of the float estimator.

It can be shown that in the one-dimensional case $\bar{a}$ will always lie in-between $\hat{a}$ and $\check{a}$, so that $|\hat{a} - \check{a}| \geq |\hat{a} - \bar{a}|$, which means that the BIE and the fixed estimator are pulled in the same direction in the one-dimensional case. This is shown in Figure 3 for different precisions (i.e. for different k). The ambiguity residuals are defined as $\check{\epsilon} = \hat{a} - \check{a}$ and $\bar{\epsilon} = \hat{a} - \bar{a}$. This Figure shows that indeed the relations in (9) hold.

Figure 4 shows the probability that the baseline estimators will be within a certain interval 2ϵ that is centered at the true baseline b, again for different precisions. It can be seen that for high success rates indeed the relation $P[\check{b} \in E_b] \geq P[\hat{b} \in E_b]$

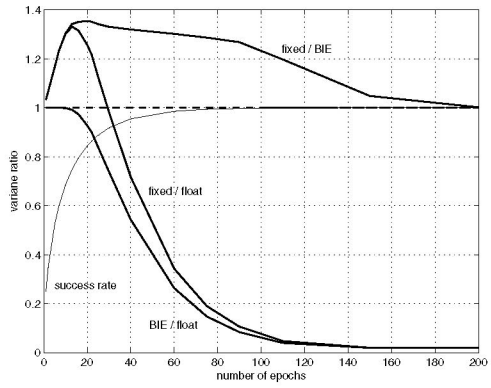


Fig. 2 Variance ratios of: BIE and float estimator; BIE and fixed estimator; fixed and float estimator. Success rate as function of the number of epochs is also shown.

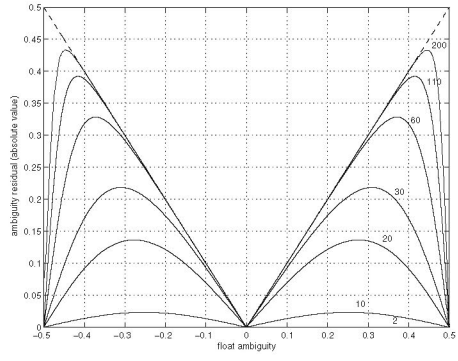


Fig. 3 Absolute values of ambiguity residuals for fixed (dashed) and BIE (solid) estimators for different number of epochs.

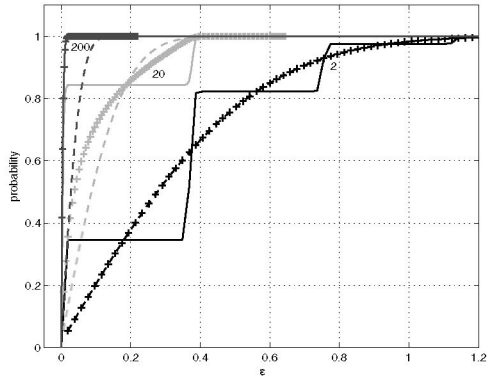


Fig. 4 Probabilities $P[|\hat{b}-b| \leq \epsilon]$ (dashed), $P[|\check{b}-b| \leq \epsilon]$ (solid), $P[|\bar{b}-b| \leq \epsilon]$ (+-signs) for different number of epochs.

k	$P(\check{a}=a)$	P_1	P_2	P_3
1	0.2490	0.5571	0.6049	0.4429
2	0.3465	0.5828	0.5871	0.4172
5	0.5221	0.6349	0.6419	0.3652
10	0.6846	0.7071	0.7163	0.2961
20	0.8436	0.8196	0.8257	0.2395
30	0.9180	0.8847	0.8867	0.3006
60	0.9860	0.9368	0.9368	0.4575
75	0.9940	0.9393	0.9393	0.4804
110	0.9991	0.9332	0.9332	0.4965
200	1.0000	0.9129	0.9129	0.5040

Table 1 Probabilities $P_1 = P(|\check{b}-b| \leq |\hat{b}-b|)$, $P_2 = P(|\bar{b}-b| \leq |\hat{b}-b|)$, and $P_3 = P(|\bar{b}-b| \leq |\check{b}-b|)$.

is true, with E_b a convex region centered at b, which means that the probability that the fixed baseline will be closer to the true but unknown baseline is larger than that of the float baseline, cf. (Teunissen, 2003b). However, for lower success rates some probability mass for $\check{b}$ can be located far from b because of the multi-modal distribution. Ideally, the probability should be high for small ϵ and reach its maximum as soon as possible. For lower success rates, the float and fixed estimators will only fulfill one of these conditions. The probability for the BIE estimator always falls more or less in-between those of the float and fixed estimators, or is equal to one of these probabilities.

The probabilities shown in Figure 4 are determined by counting the number of solutions that fall within a certain interval, but it is also interesting to compare the estimators on a sample by sample basis. In order to do so, one could determine for each sample which of the three estimators is closest to the true b, and then count for each estimator how often it was better than the other estimators.

In Table 1 the probabilities $P(|\check{b}-b| \leq |\hat{b}-b|)$, $P(|\bar{b}-b| \leq |\hat{b}-b|)$, and $P(|\bar{b}-b| \leq |\check{b}-b|)$ are given for different number of epochs. Note that if the ambiguities are fixed correctly, that does not automatically imply that $\check{b}$ is better than $\hat{b}$ because of the probability distribution of $\hat{b}$. It follows that the probability that $\bar{b}$ is better than the corresponding $\hat{b}$ is larger or equal to the probability that $\check{b}$ is better than $\hat{b}$. That is because the ambiguity residuals that are used to compute the fixed and BIE baseline estimator, see eqs.(3) and (6), have the same sign, and $|\hat{a}-\bar{a}| \leq |\hat{a}-\check{a}|$ as was shown in Figure 3. If the $\hat{b}$ is already close to the true solution, it is possible that $\bar{b}$ is closer to b, but that $\check{b}$ is pulled 'over' the true solution so that it has the opposite sign as $\hat{b}$ and the distance to b is larger. Note

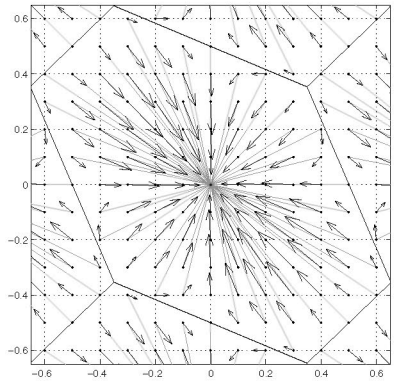


Fig. 5 For different float ambiguity vectors the corresponding BIE and fixed estimator are determined. The arrows point from $\hat{a}$ to corresponding $\bar{a}$. $\sigma_p = 0.8$m, $\sigma_\phi = 8$mm.

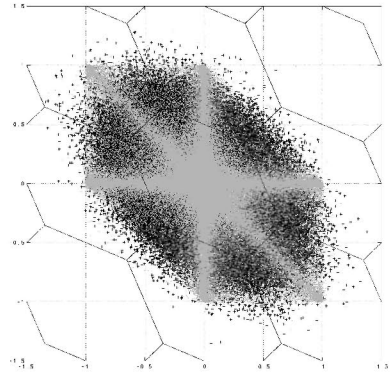


Fig. 6 Scatter plot of float ambiguities (black) and BIE ambiguity estimates (grey). $\boldsymbol{\sigma_p} = 0.6$m, $\boldsymbol{\sigma_\phi} = 6$mm.

that this does not necessarily hold true for the higher dimensional case ($n > 1$). For $k > 1$ it turns out that if $\check{b}$ is better than $\hat{b}$, then also $\bar{b}$ will better than $\hat{b}$.

Comparison in the two-dimensional case

A similar approach was followed to generate samples for the geometry-free dual-frequency GPS model for one satellite-pair. The double difference standard deviations were chosen such that success rates between 0.3 and 0.99 were obtained, see Table 2.

Figure 5 shows an example of the BIE and fixed ambiguity estimates that correspond to certain float ambiguity vectors. The hexagons are the pull-in regions; all float solutions that fall in a specific pull-in region are fixed to the corresponding integer grid point in the center of this region in case of integer least-squares. It can be seen that the BIE estimator is also pulled in approximately the same direction, but the ambiguity residuals are always much smaller for this example. If the float solution falls close to an integer grid point, the BIE estimator is pulled in that direction. On the other hand, if the float solution falls close to the boundary of a pull-in region, which means that the distance to at least two integer grid points is approximately the same, the ambiguity residual is small and the BIE estimator is pulled in another direction.

Figure 6 shows a scatter plot of simulated $\hat{a}$ and corresponding $\bar{a}$. For higher precisions, the 'star'-shape of the distribution of $\bar{a}$ becomes even stronger; for lower precisions the distribution becomes more ellipse-shaped like that of $\hat{a}$.

The distribution of the three baseline estimators is shown for different precisions. The probabilities

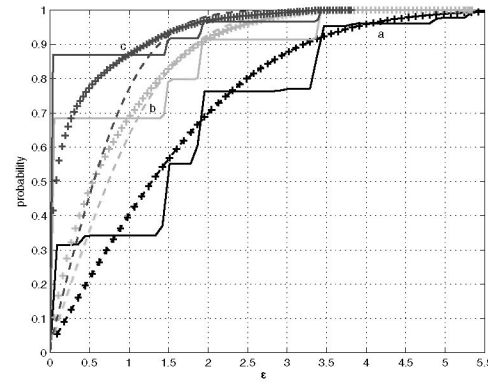


Fig. 7 $P[|\hat{b} - b| \leq \epsilon]$ (dashed), $P[|\check{b} - b| \leq \epsilon]$ (solid), $P[|\bar{b} - b| \leq \epsilon]$ (+-signs). a) $\sigma_p = 1.4$m, $\sigma_\phi = 14$mm; b) $\sigma_p = 0.8$m, $\sigma_\phi = 8$mm; c) $\sigma_p = 0.6$m, $\sigma_\phi = 6$mm.

that the baseline estimators will be within a certain interval 2ϵ that is centered at the true baseline b is shown in Figure 7.

Table 2 shows the probabilities $P(|\check{b}-b| \leq |\hat{b}-b|)$, $P(|\bar{b} - b| \leq |\hat{b} - b|)$, and $P(|\bar{b} - b| \leq |\check{b} - b|)$. This Table only shows the probabilities that the estimators will perform better than the other estimators. However, this does not say how much better, or how much worse in the instances that they do not perform better. Therefore, one should also consider the probabilities as shown in Figure 7. From the Table it follows namely that $P(|\check{b} - b| \leq |\hat{b} - b|)$ is larger than $P(|\bar{b} - b| \leq |\hat{b} - b|)$. On the other hand, Figure 7 shows that if $\check{b}$ is not close to b, it will immediately be far from the true solution (because of the stepwise function), whereas for the BIE estimator that is not necessarily the case.

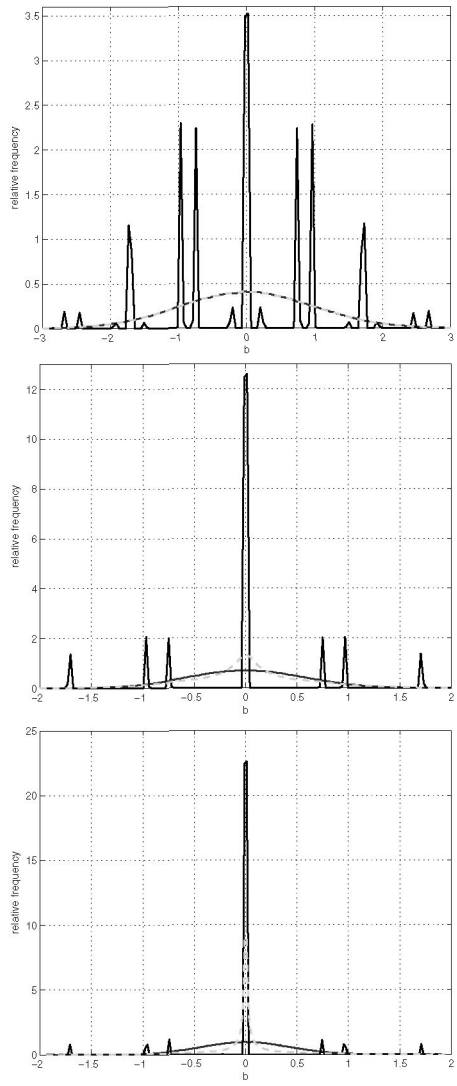


Fig. 8 Examples of distributions of the different baseline estimators: $\hat{b}$ (solid), $\check{b}$ (solid, multi-modal), $\bar{b}$ (dashed). Top: $\sigma_p = 1.4$m, $\sigma_\phi = 14$mm; Center: $\sigma_p = 0.8$m, $\sigma_\phi = 8$mm; Bottom: $\sigma_p = 0.6$m, $\sigma_\phi = 6$mm.

σ	P_s	P_1	P_2	P_3
1.4	0.3119	0.5256	0.5209	0.4744
0.8	0.6749	0.7237	0.7076	0.2827
0.6	0.8591	0.8745	0.8593	0.2050
0.4	0.9853	0.9834	0.9812	0.4336

Table 2 Probabilities $P_s = P(\check{a} = a)$, $P_1 = P(|\check{b} - b| \leq |\hat{b} - b|)$, $P_2 = P(|\bar{b} - b| \leq |\hat{b} - b|)$, and $P_3 = P(|\bar{b} - b| \leq |\check{b} - b|)$. Double difference code and phase standard deviations respectively: $\sigma_p = \sigma$m and $\sigma_\phi = \sigma$cm.

4 Concluding remarks

The key to high-precision GNSS positioning is to make use of the constraint that the unknown double difference ambiguities are integer-valued. In practice therefore the ambiguities are fixed using an appropriate ambiguity resolution method. Only if the probability of correct integer estimation, referred to as the success rate, is very high the fixed ambiguities are then used to find the corresponding fixed baseline solution. If, on the other hand, the success rate is not considered high enough, one has to stick with the float solution and collect more data before a fixed solution can be obtained.

In this paper it has been shown, that the BIE estimator might be useful to circumvent this problem. This estimator is always 'best' in the sense that it minimizes the mean squared errors and at the same it will give almost the same results as the fixed estimor when the precision is very high, and it will give almost the same results as the float solution when the precision is very low.

The next steps are to implement BIE estimation such that it can be used for real GPS applications and to see how it performs then.

References

Betti, B., Crespi, M., and Sanso, F. (1993). A geometric illustration of ambiguity resolution in GPS theory and a bayesian approach. *Manuscripta Geodaetica*, 18:317–330.

Gundlich, B. (2002). *Statistische Untersuchung ganzzahliger und reellwertiger unbekannter Parameter im GPS-Modell.* Ph.D. thesis, DGK, Reihe C, no. 549, Muenchen.

Gundlich, B. and Koch, K.-R. (2002). Confi dence regions for GPS baselines by Bayesian statistics. *Journal of Geodesy*, 76:55–62.

Teunissen, P. J. G. (1993). Least squares estimation of the integer GPS ambiguities. *Invited lecutre, Section IV Theory and Methodology, IAG General Meeting, Beijing.*

Teunissen, P. J. G. (1999). An optimality property of the integer least-squares estimator. *Journal of Geodesy*, 73:587–593.

Teunissen, P. J. G. (2003a). GNSS Best Integer Equivariant Estimation. *Presented at IUGG 2003, session G.04, June 30 - July 11, Sapporo, Japan.*

Teunissen, P. J. G. (2003b). Theory of integer equivariant estimation with application to GNSS. *Journal of Geodesy*, 77:402–410.

Parametric versus non-parametric methods for optimal weighted averaging of noisy data sets

C. Kotsakis
Department of Geomatics Engineering, University of Calgary
2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4, Email: ckotsaki@ucalgary.ca

I.N. Tziavos
Department of Geodesy and Surveying, Aristotle University of Thessaloniki
University Box 440, GR-54124, Thessaloniki, Greece, Email: tziavos@olimpia.topo.auth.gr

Abstract. Averaging is a simple, yet very effective, technique that can be used for the fusion of repeated data sets and/or multiple estimates of a single parameter vector. With the increasing number of geodetic data sensors and the availability of numerous geodetic signal realizations from various satellite missions, averaging offers an important tool which can significantly reduce the data load while improving the quality of the recovered signal information. A preliminary study of weighted averaging methods for multivariate data ensembles is given in this paper, along with a novel approach for optimal weight determination directly from the available data.

1. Introduction

The problem of *optimal weighted averaging* of data sets that correspond to multiple noisy recordings of a given signal is studied in this paper. Being one of the earliest least-squares applications, the method of averaging provides the simplest and often most efficient technique for noise filtering when several measurements of the same quantity are available. It can also serve as a trend determination and/or removal tool from different random field realizations, as well as a combination strategy in cases where various estimates of an unknown signal (or vector) need to be merged into a single optimal solution. Examples from the use of averaging techniques in geodesy can be found in the analysis and stacking of repeat-track altimetric data (Knudsen, 1993; Sailor, 1994), in the combination procedures employed by IGS to produce precise clock and satellite orbit solutions from the individual submissions of its Analysis Centers (Beutler et al., 1995; Kouba et al., 1995), and in the recovery of topographic gradients from stacked repeat-pass SAR interferograms without phase unwrapping (Sandwell and Sichoix, 2000). We should also mention another important geodetic paradigm whose strong dependence on the proper application of an averaging algorithm has been often overlooked, namely the integrated statistical processing of orthometric, ellipsoidal and geoidal heights for vertical datum and gravimetric geoid refinement studies (Kotsakis, 2003).

Generalizing the traditional estimation problem in statistics, where the expectation of a single random variable is approximated by the sample mean of a set of its observed values, *signal* or *vector averaging* can be used for the analysis and fusion of multivariate data ensembles. The issue of proper weighting of the different ensembles becomes especially important in this case, since a *single* weight factor may not be able to account for: (i) the noise correlation within each data set, or (ii) the noise cross-correlation between different data sets. This can be very critical in applications where the errors affecting the repeated realizations of a geodetic signal are significantly correlated at various spatial and/or temporal scales (e.g., atmospheric effects). Furthermore, in many cases the statistical characteristics of the data errors are only poorly known, thus introducing additional difficulties for choosing the appropriate data weights.

In this paper, we present a preliminary treatment of the weighted averaging problem for multivariate data ensembles, with the emphasis put on two different approaches for the optimal determination of the ensemble weights. The first approach, which is useful for theoretical or simulation studies, assumes the complete knowledge of the second-order statistics for the data noise. Some examples for the error behaviour of the corresponding optimal averaging estimator are presented to demonstrate the importance of proper weighting when mixing data sets with varying accuracy levels. The second approach is oriented towards more practical situations and it is based on an 'empirical' estimation criterion that allows the determination of

the optimal ensemble weights directly from the available data.

2. Problem formulation

Let us assume that a set of N multivariate noisy observations (ensembles) of a given signal is available. For notational simplicity these ensembles are represented as column vectors $\mathbf{x}_i$ in an m-dimensional *data space* $\Re^m$, where m is the number of data values in each recorded signal realization. Note that unless we are dealing with 1D problems, this formulation requires the choice of a one-to-one mapping between the physical signal description (e.g., the raster scan of a digital image) and the components of the data vectors $\mathbf{x}_i$ (ensemble vectorization).

Assuming additive noise in each data set, the following observation equations can now be formed:

$$\mathbf{x}_i = \boldsymbol{\mu} + \mathbf{v}_i \ , \quad i = 1, 2, ..., N \tag{1}$$

where the vector $\boldsymbol{\mu}$ contains the true values of the underlying signal. The random error vectors $\mathbf{v}_i$ have zero mean and they are statistically described in terms of their auto- and cross-covariance matrices

$$\begin{aligned} \mathbf{C}_{ii} &= E\{\mathbf{v}_i \mathbf{v}_i^{\mathrm{T}}\}, \\ \mathbf{C}_{ij} &= E\{\mathbf{v}_i \mathbf{v}_j^{\mathrm{T}}\} \qquad i, j = 1, 2, ..., N \end{aligned} \tag{2}$$

where $E\{\cdot\}$ is the mathematical expectation operator. Note that all vector(-ized) ensembles are assumed to have the same dimensions and they correspond to the same spatial (or temporal) sampling configuration.

A linear estimator of the unknown signal from its repeated realizations has the general form

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{N} \mathbf{W}_i \mathbf{x}_i \tag{3}$$

The last formula can easily be identified as a weighted sum of the ensembles $\{\mathbf{x}_i\}$, with $\{\mathbf{W}_i\}$ being appropriate weight matrices that need to be determined in some optimal sense. Due to the limited extent of this paper, we will restrict our study to the case where the weight matrices have the simple diagonal form

$$\mathbf{W}_i = w_i \mathbf{I} \tag{4}$$

where $\mathbf{I}$ denotes the $m \times m$ unit matrix. In this way, the linear estimator of Eq. (3) is simplified to

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{N} w_i \mathbf{x}_i \tag{5}$$

where a single weight coefficient w_i is now assigned to each data set. The treatment of the more general case of weighted averaging, according to Eq. (3), will soon be published elsewhere.

Remark: The previous formulation is also suited for problems where the data sets $\mathbf{x}_i$ are not obtained by direct repeated observations of the same field, but they correspond to different a-priori estimates or solutions of an unknown vector quantity. The various ensembles $\{\mathbf{x}_i\}$ then need to be optimally fused into a unique combined solution $\hat{\boldsymbol{\mu}}$ through a weighted averaging procedure according to Eq. (3) or Eq. (5). In such cases, the covariance (CV) matrices given in Eq. (2) represent the accuracy of each individual vector estimate and they are usually determined from separate data adjustment procedures. The cross-CV matrices $\mathbf{C}_{ij}$ should normally vanish if the prior estimates for $\boldsymbol{\mu}$ have been obtained independently from each other. A typical example of this type of problem can be found in the combination strategies implemented by IGS for the determination of precise orbit and clock solutions (Beutler et al., 1995).

3. Rigorous weight determination

The computation of the weight factors in Eq. (5) should be based on a well-defined criterion that optimizes the output signal estimate. In this section, a rigorous *mean square error* (MSE) criterion is adopted for the optimal determination of the unknown weights. Denoting by $\mathbf{e}$ the actual error of the linear estimator in Eq. (5), i.e.

$$\mathbf{e} = \hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \tag{6}$$

the corresponding MSE is a certain functional that quantifies in an average sense the magnitude of the above estimation error. If we use the standard Euclidean vector norm, then it can be shown that the MSE can be expressed by the following form:

$$\begin{aligned} E\{\|\mathbf{e}\|^2\} &= E\{\mathbf{e}^{\mathrm{T}}\mathbf{e}\} \\ &= \mathbf{w}^{\mathrm{T}} \tilde{\mathbf{C}} \mathbf{w} + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} \left[\left(\sum_{i=1}^{N} w_i \right) - 1 \right]^2 \end{aligned} \tag{7}$$

where the vector $\mathbf{w}$ contains the values of the scalar weights (w_i), and the matrix $\tilde{\mathbf{C}}$ is defined by the formula

$$\tilde{\mathbf{C}}:\ [\tilde{C}_{ij}] = trace\,\mathbf{C}_{ij} \qquad i,j=1,2,...,N \qquad (8)$$

If we further impose the condition that the result of Eq. (5) should lead to an *unbiased* signal estimate, then we are faced with the following constrained optimization problem:

$$\mathbf{w}^{\mathrm{T}}\tilde{\mathbf{C}}\mathbf{w} = minimum, \text{ subject to } \sum_{i=1}^{N} w_i = 1 \qquad (9)$$

Using the method of Lagrange multipliers, a unique weight solution can be obtained as follows:

$$\hat{\mathbf{w}} = \frac{1}{\mathbf{s}^{\mathrm{T}}\tilde{\mathbf{C}}^{-1}\mathbf{s}}\tilde{\mathbf{C}}^{-1}\mathbf{s} \qquad (10)$$

where $\mathbf{s}$ is an auxiliary vector whose elements are all ones. The minimum MSE value that corresponds to the above optimal weight vector is

$$MSE_{\min} = (\mathbf{s}^{\mathrm{T}}\tilde{\mathbf{C}}^{-1}\mathbf{s})^{-1} \qquad (11)$$

For comparison, it is interesting to give also the value that the MSE takes when a *uniform* weighting scheme is followed (i.e. $w_i = 1/N$). In such a case we have

$$MSE\ (\text{for equal weights}) = (\mathbf{s}^{\mathrm{T}}\tilde{\mathbf{C}}\mathbf{s})/N^2 \qquad (12)$$

Note that the last expression is generally sub-optimal and it decays much slower, for increasing sample size N, than the optimal error value given in Eq. (11); see also Sect. 4.

In the special case where the matrix $\tilde{\mathbf{C}}$ is diagonal (i.e. the noise cross-correlation between different data sets is zero, $\mathbf{C}_{ij} = \mathbf{0}$), the optimal weights associated with the ensembles $\mathbf{x}_i$ are given by the formula

$$\hat{w}_i = \left[\frac{1}{\sum_{k=1}^{N}\left(1/trace\,\mathbf{C}_{kk}\right)}\right]\frac{1}{trace\,\mathbf{C}_{ii}} \qquad (13)$$

whereas the resultant minimum MSE value is

$$MSE_{\min} = \frac{1}{\sum_{k=1}^{N}\left(1/trace\,\mathbf{C}_{kk}\right)} \qquad (14)$$

Again, for comparison purposes, the MSE value obtained by using equal weight coefficients (in uncorrelated data ensembles) is given below

$$MSE\ (\text{for equal weights}) = \frac{\sum_{k=1}^{N} trace\,\mathbf{C}_{kk}}{N^2} \qquad (15)$$

Remark: The MSE of the weighted average $\hat{\mu}$, defined by Eq. (7), is insensitive to the noise correlation that may exist *within* each data set $\mathbf{x}_i$. That is because the off-diagonal elements of the CV matrices $\mathbf{C}_{ii}$, which contain the statistical information for the noise correlation within every ensemble, do not enter into the definition of the matrix $\tilde{\mathbf{C}}$ according to Eq. (8). As a result, an averaging scheme that employs only scalar weighting coefficients may not be suitable for data sets that have highly correlated errors, since the individual weights obtained by the MSE minimization in Eq. (9) cannot reflect this type of information.

The *cross-ensemble* noise correlation, on the other hand, is not totally ignored in the optimal weights given in Eq. (10). The MSE definition encompasses the traces of all cross-CV matrices $\mathbf{C}_{ij}$, and thus the weight optimization according to Eq. (9) takes into account the noise correlation at every data point between different ensembles. Obviously, any statistical information on the noise cross-correlation between different ensembles and different data points is excluded from the optimal weights of Eq. (10).

4. Error analysis for the weighted average: an example

A simple example is presented in this section to demonstrate the importance of proper weighting in data averaging procedures. A series of 40 different ensembles, all corresponding to the same gravity signal profile, is considered. The ensembles are assumed to be affected by additive stationary noise whose standard deviation varies among them according to the simulated values shown in Figure 1.

Assuming also that the data noise is uncorrelated between the different ensembles, we can use Eq. (14) to plot the cumulative MSE of the optimal weighted average $\hat{\mu}$, as the number of the processed ensembles (N) increases from 2 up to 40. In this case, we will have that

$$trace\,\mathbf{C}_{kk} = m\sigma_k^2 \qquad (16)$$

where σ_k is the noise standard deviation of the k-th ensemble (see Figure 1), and m denotes the constant

number of data points within each ensemble. In a similar way, Eq. (15) can be used to determine the cumulative MSE of the unweighted (or, more precisely, uniformly weighted) signal average $\hat{\mu}$ as a function of the number of the averaged ensembles.

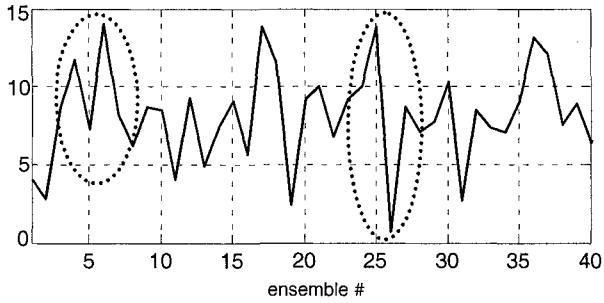


Figure 1. Simulated values (in μgal) for the noise standard deviation in the various gravity signal ensembles.

The results for both cases are shown in Figure 2, which clearly reveals the problems of uniform weighting for data sets with uneven accuracy levels. It is interesting to point out the two regions that are circled in the plots of Figure 2. The first (upper left) region shows that the accuracy of the unweighted average can actually worsen whilst more data sets are used for its computation! As seen in Figure 1, the noise level in the first two data sets is much smaller than in the next few ensembles. This affects the MSE of the unweighted average $\hat{\mu}$ accordingly when the more noisy ensembles enter into the averaging procedure. On the other hand, if the varying noise level in the different ensembles is properly taken into account through an optimal weighting scheme (see Sect. 3), then the accuracy of the averaged estimate will be consistently improving as more data sets are processed. Nonetheless, the MSE of the unweighted average seems to 'stabilize' and starting to decrease constantly when more precise data sets enter into the averaging procedure, although its decay rate is still slower than in the optimal weighting case.

In the second circled region shown in Figure 2, we notice the drastic MSE reduction in the optimal weighted average when a very accurate data set is used to update $\hat{\mu}$ (see also the corresponding circled area in Figure 1 where the decrease in the noise level between the 25th and the 26th ensemble is indicated). In the case of the uniformly weighted average (dotted line) no such considerable improvement occurs, since the equal weighting of all data sets cannot offer any benefit from the use of a significantly more accurate signal ensemble.

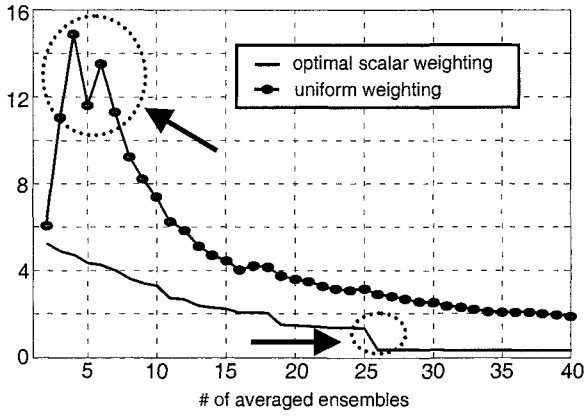


Figure 2. Cumulative MSE (in μgal^2) for the weighted signal average using: (i) optimal weights, and (ii) equal weights.

The effect of the noise correlation (between different signal ensembles) on the weighted average $\hat{\mu}$

The different gravity signal ensembles are again presumed to be affected by additive stationary noise whose standard deviation varies according to the values of Figure 1. In addition, it is assumed that a pointwise cross-ensemble correlation of the data noise exists, for all pairs of the available ensembles, with a constant correlation coefficient value of ρ= 0.7. This is a purely theoretical scenario, which nevertheless allows us to easily simulate a fully populated matrix $\tilde{\mathbf{C}}$ that is needed for a rigorous error analysis of the weighted average $\hat{\mu}$ according to Eqs. (11) and (12). In this case, the off-diagonal elements of $\tilde{\mathbf{C}}$ will be given by the simple formula

$$[\tilde{C}_{ij}] = trace\mathbf{C}_{ij} = m\rho\sigma_i\sigma_j \qquad (17)$$

where σ_i and σ_j are the noise standard deviations of the i-th and j-th ensemble, respectively. Three different schemes for computing the weighted average $\hat{\mu}$ are now considered, namely using: (i) the optimal weights according to Eq. (10), (ii) equal weights, and (iii) the weights obtained from Eq. (13) which take into account only the different noise levels in the various data sets (i.e. they neglect the cross-ensemble noise correlation).

As before, the cumulative MSEs of the corresponding signal averages are computed and plotted against the number of the averaged ensembles (N). The results are shown in Figure 3, where we can see the 'explosive' behavior of the resulting MSE for the uniformly weighted average.

Even in the case where the weights account for the uneven noise levels among the various data sets (dashed line), the corresponding weighted average still seems to be unstable as the number of processed ensembles increases. This is due to the neglected noise cross-correlation between the different ensembles, which in practice can cause serious problems in the fusion of repeated geodetic sets (e.g., sea surface topography from single/multiple altimetry missions, gravity and gradiometry maps from single/multiple satellite missions, etc.).

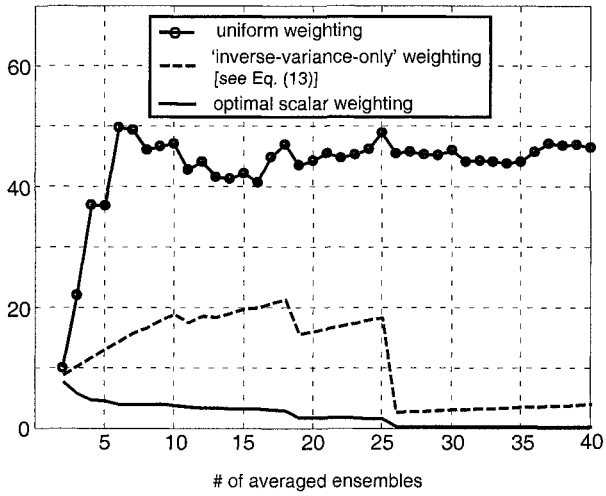


Figure 3. Cumulative MSE (in μgal²) for the weighted signal average in the case of correlated ensembles ($\mathbf{C}$ is a fully populated matrix).

Note that Figures 2 and 3 show normalized MSE values, since the results of Eqs. (11), (12), (14) and (15) have been divided by the number of data points (m) within every ensemble.

5. Empirical weight determination

The weight optimization procedure that was described in Sect. 3 requires the knowledge of the noise CV matrices for all available data sets. A more realistic approach for the determination of the ensemble weights should consider that the noise statistics are often unknown in practice (or, at least, partially known). This applies also for the case where the CV matrices in Eq. (2) represent not the statistical variability of actual measurement noise, but the accuracy of some a-priori vector estimates $\mathbf{x}_i$ which need to be merged into a single solution $\hat{\boldsymbol{\mu}}$.

In this section, an alternative optimization method is presented which results in ensemble weights that can be directly computed from the available data. It can thus be characterized as a *non-parametric* method (as opposed to the parametric approach of Sect. 3), since any a-priori assumptions for the statistical properties of the various data sets are completely avoided.

Data weighting based on a maximum-SNR criterion

The criterion that will be now used for the optimal determination of the ensemble weights is based on the maximization of an 'empirical' signal-to-noise ratio (SNR) expression. In general, the SNR that corresponds to the averaging estimator of Eq. (5) has the form

$$SNR = \frac{E_s}{E_n} = \frac{\hat{\boldsymbol{\mu}}^{\mathrm{T}}\hat{\boldsymbol{\mu}}}{E_n} \tag{18}$$

where E_s and E_n denote the Euclidean norms of the estimated signal and the data noise, respectively.

Two different methods can be followed to quantify the data noise norm. In particular, we have

Case I

$$\mathbf{e}_i = \mathbf{x}_i - \hat{\boldsymbol{\mu}}\ , \quad i=1,2,...,N \tag{19}$$

$$E_n = \sum_i w_i \left\|\mathbf{e}_i\right\|^2 = \sum_i w_i (\mathbf{x}_i - \hat{\boldsymbol{\mu}})^{\mathrm{T}}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}) \tag{20}$$

or an alternative approach,

Case II

$$\mathbf{e}_i^{(w)} = w_i \mathbf{x}_i - \hat{\boldsymbol{\mu}}\ , \quad i=1,2,...,N \tag{21}$$

$$E_n = \sum_i \left\|\mathbf{e}_i^{(w)}\right\|^2 = \sum_i (w_i\mathbf{x}_i - \hat{\boldsymbol{\mu}})^{\mathrm{T}}(w_i\mathbf{x}_i - \hat{\boldsymbol{\mu}}) \tag{22}$$

The previous approaches are considered 'empirical' in the sense that they both quantify the data noise not with respect to the true unknown signal $\boldsymbol{\mu}$, but with reference to its optimal estimate $\hat{\boldsymbol{\mu}}$. Note that *Case I* provides a straightforward generalization of the weighted least-squares error norm for the case of multivariate measurements $\{\mathbf{x}_i\}$.

Based on Eqs. (19) and (20), the SNR of the weighted signal average for *Case I* is

$$SNR_{(\mathrm{I})} = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{R}\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{D}\mathbf{s} - \mathbf{w}^{\mathrm{T}}\mathbf{R}\mathbf{w}} \tag{23}$$

where $\mathbf{R}$ is a symmetric Grammian-type matrix that consists of the Euclidean inner products $(\mathbf{x}_i, \mathbf{x}_j)$ between all possible pairs of data sets

$$\mathbf{R}: \; [R_{ij}] = \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j \qquad i, j = 1, 2, ..., N \qquad (24)$$

and **D** is a diagonal matrix obtained from **R** by setting to zero its off-diagonal elements

$$\mathbf{D}: \; [D_{ii}] = \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_i \qquad i = 1, 2, ..., N \qquad (25)$$

The maximization of the SNR expression in Eq. (23) yields a unique solution that is given by the closed formula

$$\underline{\textit{Case I}}: \qquad \hat{\mathbf{w}} = \frac{1}{\mathbf{s}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{D} \mathbf{s}} \mathbf{R}^{-1} \mathbf{D} \mathbf{s} \qquad (26)$$

As it can be seen from the last equation, the optimal weight vector can now be computed directly from the available data, without any other knowledge of their statistical properties.

If the second formulation for the data noise norm E_n is used (*Case II*), then the corresponding SNR of the weighted signal average takes the form

$$SNR_{(\mathrm{II})} = \frac{\mathbf{w}^{\mathrm{T}} \mathbf{R} \mathbf{w}}{\mathbf{w}^{\mathrm{T}} \mathbf{D} \mathbf{w} - \mathbf{w}^{\mathrm{T}} \mathbf{R} \mathbf{w}} \qquad (27)$$

which is slightly different from the previous SNR expression in Eq. (23). Its maximization leads to a *generalized eigenvalue problem*, as follows:

$$\underline{\textit{Case II}}: \qquad \mathbf{R}\mathbf{w} = \lambda \mathbf{D}\mathbf{w} \qquad (28)$$

In this case, the optimal weight vector $\hat{\mathbf{w}}$ is just the eigenvector corresponding to the maximum eigenvalue λ_{max} in Eq. (28); more mathematical details and proofs will be given in an upcoming journal publication. Note that generalized eigenvalue problems can be easily converted to standard eigenvalue problems (i.e. $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$) through simple linear transformations; see Hairville (1997, pp. 562-564). Most high-end scientific computing packages, such as Matlab™, contain also built-in routines for solving directly this type of algebraic problems.

Remark: Both of the 'empirically optimal' weight solutions described above will produce unbiased signal estimates when used in the averaging estimator of Eq. (5). That is because the sum of their values $\{w_i\}$ is, or it can be easily made after a simple global rescaling, equal to one. Note that this does not exclude cases where the optimal weight solution contains a mixture of both positive and negative values, as long as their total sum is still equal to one! Actually, such cases will arise often when the data noise shows significant cross-correlation among the different signal ensembles.

6. Conclusions

The practice of averaging multiple realizations of a given signal, either in the form of repeatedly observed ensembles or as a-priori estimates obtained from different individual procedures, is a valuable tool in geodetic data analysis. In this paper we have briefly presented a general treatment of the weighted averaging problem for multivariate data sets. Our main focus has been on: (i) the optimal determination of the ensemble weights (with and without the knowledge of the data noise statistics), and (ii) the error behaviour of the signal average for different weighting schemes. In terms of future work we plan to present a more detailed version of the study outlined herein that will include all of the omitted mathematical details, as well as more sophisticated weighting schemes for data averaging such as the use of fully populated weight matrices $\mathbf{W}_i$ and the simultaneous bias estimation/removal from the data sets $\{\mathbf{x}_i\}$.

References

Beutler G, Kouba J, Springer T (1995) Combining the orbits of the IGS Analysis Centers. *Bulletin Géodésique*, 69:200-222.

Hairville DA (1997) *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, New York.

Knudsen P (1993) Altimetry for geodesy and oceanography. In *Lecture notes for the NKG-Autumn School 1992* (J. Kakkuri, ed.): "Geodesy and Geophysics", Report of the Finnish Geodetic Institute, No. 115, pp. 87-130, Helsinki.

Kouba J, Mireault Y, Lahaye F (1995) IGS Orbit/Clock Combination and Evaluation (Appendix I of the Analysis Coordinator Report), *International GPS Service for Geodynamics (IGS) 1994 Annual Report*, pp. 70-94.

Kotsakis C (2003) Some remarks on the integrated statistical processing of GPS, levelling and geoid data (in preparation).

Sailor RV (1994) Signal Processing Techniques. In *Geoid and its Geophysical Interpretations* (P. Vanicek and N.T. Christou, eds.), pp. 147-186, CRC Press.

Sandwell DT, Sichoix L (2000) Topographic phase recovery from stacked ERS interferometry and a low resolution digital elevation model. *Journal of Geophysical Research*, vol. 105, no. B12, pp. 28,211-28,222.

Estimation of Variance Components Through a Combined Adjustment of GPS, Geoid and Levelling Data

G. Fotopoulos, C. Kotsakis, and M.G. Sideris
Department of Geomatics Engineering, University of Calgary
2500 University Drive N.W., Calgary, Alberta, Canada, T2N 1N4, Email: gfotopou@ucalgary.ca

Abstract. The method of GPS-levelling for obtaining orthometric heights is not a new concept. In fact, many studies have proven its usefulness and the question of whether GPS-levelling can provide a viable alternative to traditional techniques is no longer an issue. An important question, however, that has yet to be satisfactorily solved is, *'What accuracy level can be achieved using this approach?'* Over the past decade, numerous advances have been made which have placed us in a position where we can begin to address the issue with more confidence, namely (i) improved mathematical models/techniques for dealing with GPS and geoid data, (ii) increased data availability for gravimetric geoid models, and (iii) improved data processing capabilities. In this paper a statistical approach for estimating the variance components of heterogeneous groups of observations is used in the combined adjustment of GPS, geoid and levelling data. Specifically, the iterative minimum norm quadratic unbiased estimation algorithm is employed to determine the individual variance components for each of the three height types. The challenges encountered when implementing this well-known algorithm in practice with real data are discussed. The analysis provides some indication into the practicality and effectiveness of estimating variance components in mixed vertical networks. Notably, the estimation of realistic variance components provides us with important insight regarding the GPS-levelling problem in addition to other uses of combined GPS, geoid and levelling data, such as assessing the accuracy of a gravimetric geoid model.

1 Introduction

The reliable combined adjustment of GPS (h), orthometric (H) and geoid (N) height data depends on two main factors, namely (i) the suitability of a parametric model for the systematic effects and biases and (ii) the correctness of the stochastic model for the observational noise. The former has been a topic of interest for researchers and practitioners alike in light of its benefits for GPS-based determination of orthometric heights in a local vertical datum. Although the use of a parametric model designed to absorb the systematic effects and datum inconsistencies inherent among the GPS, geoid and levelling height data is an interesting topic on its own, it will not be dwelled on extensively herein. Instead, the issue of stochastic modelling for observations in the mixed adjustment of heterogeneous height data will be the main focus of this paper. It should, however, be emphasized that both aforementioned factors are interrelated and need to be addressed in order to rigorously and reliably combine the different types of height data. Of course, both the modelling of systematic effects and random errors presupposes the absence of gross errors/blunders in the observational data. Furthermore, improper stochastic modelling may lead to certain deviations in the results, the very effect we try to model through parameterization of the height misclosures ($h - H - N$).

The motivation for addressing this issue is quite clear given the vast applications requiring optimally combined height data, including, but not limited to:

- modernize regional vertical datums
- unify national regional datums between neighbouring countries
- transform between various height data sets (incl. vertical reference systems)
- refine existing gravimetric geoid models

The chosen approach presented herein for testing/improving the stochastic model is the well known statistical tool of variance component estimation (VCE). Many different algorithms for VCE have been studied with regards to geodetic data analysis and a comprehensive literature review is given in Grafarend (1985). Recent geodetic

applications include the statistical analysis of VLBI data, GPS observations, deformation monitoring schemes and simulated gravity field models from upcoming satellite missions such as GOCE. Despite these advances, the implementation of VCE for combined data types, more specifically GPS, orthometric and geoid heights, has not been suitably addressed in the geodetic literature. A properly applied VCE technique coupled with appropriate stochastic modelling will allow for the optimal combination of these heterogeneous data leading to the inclusive goal of more realistic accuracy measures for height-related applications.

2 Problem Description

Given a network of points with co-located GPS, orthometric and geoid height data, a combined adjustment can be performed. We start with the general linear model (see Kotsakis and Sideris, 1999 for details),

$$\mathbf{l} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{v} \tag{1a}$$

$$E\{\mathbf{v}\} = \mathbf{0} \tag{1b}$$

$$E\{\mathbf{v}\mathbf{v}^{\mathrm{T}}\} = \mathbf{C}_{\mathbf{v}} = \sigma^2\mathbf{Q}_{\mathbf{v}} \tag{1c}$$

where the vector of observations $\mathbf{l}$ is composed of the height 'misclosure' at each GPS/levelling benchmark as follows

$$l_i = h_i - H_i - N_i \tag{2}$$

$E\{\cdot\}$ is the mathematical expectation operator, $\mathbf{A}$ is the design matrix which depends on the parametric model type (see Sect. 5.4), $\mathbf{B}$ is the block-structured matrix $\mathbf{B} = [\mathbf{I} \ -\mathbf{I} \ -\mathbf{I}]$, where each $\mathbf{I}$ is an $m \times m$ unit matrix (m is the number of observation equations), $\mathbf{x}$ is a vector containing the unknown parameters corresponding to the selected parametric model and $\mathbf{v}$ is a vector of random errors, with zero mean (Eq. 1b), described by the following formula

$$\mathbf{v} = \left[\mathbf{v}_{\mathrm{h}}^{\mathrm{T}} \quad \mathbf{v}_{\mathrm{H}}^{\mathrm{T}} \quad \mathbf{v}_{\mathrm{N}}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{3}$$

where $\mathbf{v}_{(\cdot)}$ is an $m \times 1$ vector of random errors for each of the h, H, N data types. The corresponding CV matrix is described by Eq. (1c), where the positive-definite cofactor matrix $\mathbf{Q}_{\mathbf{v}}$ is scaled by the variance factor σ^2. For the case of heterogeneous disjunctive observations, a block-diagonal CV model is used

$$\mathbf{C}_{\mathbf{v}} = \begin{bmatrix} \mathbf{C}_{\mathbf{h}} & 0 & 0 \\ 0 & \mathbf{C}_{\mathbf{H}} & 0 \\ 0 & 0 & \mathbf{C}_{\mathbf{N}} \end{bmatrix} \tag{4}$$

where $\mathbf{C}_{(\cdot)}$ is the CV matrix for each of the height types. An additive covariance matrix model for the observations is given by:

$$\mathbf{C}_{\mathbf{l}} = \mathbf{B}\mathbf{C}_{\mathbf{v}}\mathbf{B}^{\mathrm{T}} = \sigma_h^2\,\mathbf{Q}_{\mathbf{h}} + \sigma_H^2\,\mathbf{Q}_{\mathbf{H}} + \sigma_N^2\,\mathbf{Q}_{\mathbf{N}} \tag{5}$$

where $\mathbf{Q}_{\mathbf{h}}$, $\mathbf{Q}_{\mathbf{H}}$, and $\mathbf{Q}_{\mathbf{N}}$ are known positive-definite cofactor matrices for the GPS, orthometric and geoid height data respectively. The unknown variance components are σ_h^2, σ_H^2, σ_N^2.

The problem, therefore, is to solve for the unknown parameters of the parametric model, $\mathbf{x}$, **and** the individual variance components for each of the height data types, σ_h^2, σ_H^2, σ_N^2.

3 Variance Component Estimation

There are a number of methods available to perform VCE within the context of least-squares adjustment [Crocetto et al., 2000]. In this paper the minimum norm quadratic unbiased estimation (MINQUE) procedure is followed [Rao, 1971]. It has been shown that many VCE approaches (i.e., best quadratic unbiased estimation, MINQUE), under the assumption of normally distributed observations, give seemingly numerically equivalent results. However, distinguishable characteristics in the formulation of the problem for each algorithm exist. Ultimately, the selection of the appropriate technique should rely on the desired estimator properties, such as translation invariance, unbiasedness, computational efficiency, etc. The rationale detailing our preference for this particular approach is not included due to space restrictions, however it can be stated that this approach works well for the posed problem and is relatively easy to implement in practice.

The selected VCE algorithm is briefly described below. Due to space restrictions, only the basic

formulas are given without a detailed account as to their derivation. The problem is to solve the following system

$$\mathbf{S}\hat{\boldsymbol{\theta}} = \mathbf{q} \tag{6}$$

where $\hat{\boldsymbol{\theta}}$ is a vector containing the unknown variance components (also indicated by $\hat{\boldsymbol{\sigma}}^2$). The composition of the matrix $\mathbf{S}$ is denoted by

$$\mathbf{S} = \begin{bmatrix} s_{hh} & s_{hH} & s_{hN} \\ s_{Hh} & s_{HH} & s_{HN} \\ s_{Nh} & s_{NH} & s_{NN} \end{bmatrix} \tag{7}$$

Each element s_{ij} in the matrix is computed from

$$s_{ij} = tr\left(\mathbf{R}\mathbf{Q}_i\mathbf{R}\mathbf{Q}_j\right), \quad i,j = h,H,N \tag{8}$$

where $tr(\cdot)$ is the trace operator, $\mathbf{Q}_{(\cdot)}$ is a positive-definite cofactor matrix for each group of observations, $\mathbf{R}$ is the matrix defined by

$$\mathbf{R} = \mathbf{C}_\mathbf{l}^{-1}[\mathbf{I} - \mathbf{A}(\mathbf{A}^\mathrm{T}\mathbf{C}_\mathbf{l}^{-1}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}\mathbf{C}_\mathbf{l}^{-1}] \tag{9}$$

where $\mathbf{A}$ is the same design matrix as in Eq. (1a) and $\mathbf{C}_\mathbf{l}$ is the CV matrix of the observations. The vector $\mathbf{q}$ contains the quadratic forms

$$\mathbf{q} = \{q_i\}, \quad q_i = \hat{\mathbf{v}}_i^\mathrm{T}\mathbf{Q}_i^{-1}\hat{\mathbf{v}}_i \tag{10}$$

where $\hat{\mathbf{v}}_i = \mathbf{Q}_i\mathbf{R}\mathbf{l}$ are the separated residuals for each group of observations and is easily derived from the mixed model equation formulation; see Kotsakis and Sideris (1999) for more details.

It is evident from the expression for the $\mathbf{R}$ matrix (Eq. 9) that an iterative procedure must be employed since the unknown variance components are embedded in $\mathbf{C}_\mathbf{l}$ (Eq. 5) that is used to compute $\mathbf{R}$. Therefore, initial values for the unknowns must be provided along with an appropriate convergence criterion. The iterative application of this algorithm is known as iterative minimum norm quadratic unbiased estimation (I-MINQUE). The general iterative scheme is shown in Figure 1 where each $\hat{\theta}_i^\alpha$ value represents a ratio computed at each iteration α and the final estimated variance components are obtained from the product of all ratios after n iterations. More information regarding the estimator properties can be found in Rao (1971); Rao and Kleffe (1988).

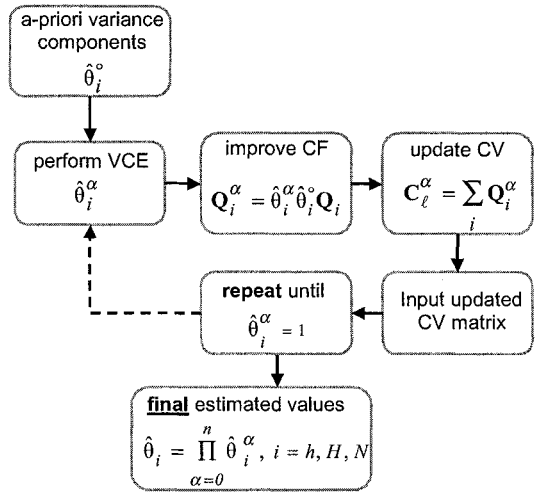


Fig. 1. Iterative variance component estimation scheme

Notes. The problem, as described in the previous section, outlines a rare characteristic for geodetic data, that of *balanced* data. Normally, when heterogeneous types of geodetic data are used in a mixed adjustment, the number of observations per each group of data is not the same. In this case, the problem is pre-designed such that all three height groups are available for each network benchmark. Thus, we can estimate the variance components from balanced data - a less demanding task, in general, than dealing with unbalanced data. An additional advantage offered by the design of this particular problem is a relatively low computational load. Only three variance components are sought. The largest matrix inversion will be on the order of the number of observations, m, which in the absolute case is equivalent to the number of network benchmarks. Thus, unlike many other VCE-related applications, where the main obstacle encountered is the high computational load, the problem here, lies in the absence of independently derived and reliable variance estimates for each height type.

4 Description of Data

The Swiss national network, consisting of 111 GPS/levelling benchmarks distributed throughout an ~343 km × 212 km region, was used for the numerical tests. At each point the observations, as given by Eq. (2), were computed (statistics: mean ~ 1.1 cm, min. –4.9 cm, max. 19 cm, RMS ~ 4 cm). Fully populated initial CV matrices were obtained for h, H and N. The original CV matrix for the ellipsoidal heights $\mathbf{Q}_h$ was extracted from the results of a commercial post-processing software package of GPS data. Typical for GPS, the output CV matrix was overly optimistic, a direct result of neglecting (even partially) correlations (temporal, spatial and physical) between GPS phases. In practice, this situation is sometimes rectified by *arbitrarily* scaling the CV matrix by some factor. The validity of this all too common practice is addressed in sect. 5.1. $\mathbf{Q}_H$ comes directly from the rigorous national adjustment of all first and second order levelling measurements. As expected, the correlation between nearby neighbouring stations is very high. Finally, $\mathbf{Q}_N$ at the GPS benchmarks was obtained by straightforward application of error propagation to the least-squares collocation geoid solution [see Marti, 2002 for more details]. Table 1 summarizes some characteristics of the 'a-priori' covariance matrices for comparison purposes.

Table 1. Initial covariance matrix characteristics

	GPS	Levelling	Geoid
condition number	146.2	2.23×10^{7}	4.50×10^{5}
average σ (cm)	0.79	0.75	1.93

The comparatively high condition number corresponding to the levelling data matrix is of particular interest. This high value may lead to numerical instability problems when the matrix needs to be inverted. Therefore, a simple ridge regression was applied to alleviate any numerical problems. Other values listed in the table include the average standard deviation (σ) computed from $\sqrt{tr(\mathbf{Q}_i)/m}$, which gives an indication of the overly optimistic a-priori levels, especially evident for the ellipsoidal heights.

Note that, where possible, the GPS/levelling observations were excluded from the computation of the gravimetric geoid to ensure independence among the cofactor matrices satisfying the assumption of disjunctive observation groups, i.e. the cross-covariance matrices for $i, j = h, H, N$ are given by:

$$E\{\mathbf{v}_i \mathbf{v}_j^T\} = \mathbf{0}, \quad for\ i \neq j \tag{11}$$

5 Numerical Examples – Case Studies

In this section, the results of four different empirical tests are described, whereby some of the key issues related to the implementation of the I-MINQUE method are analyzed in detail, demonstrating the use of VCE for practical height-related applications.

5.1 Case Study I: A-priori CV matrices

The purpose of this case study is to test the effect of different a-priori CV matrices on the final estimated variance components. The I-MINQUE scheme does not provide any guarantee of the *correctness* of the final estimated values. Therefore, mere convergence cannot be taken as a positive re-enforcer. To overcome this uncertainty, one can compute the CV matrix for $\hat{\boldsymbol{\theta}}$ using the following formula [Crocetto et al., 2000]:

$$\mathbf{C}_{\hat{\theta}} = 2\mathbf{S}^{-1} \tag{12}$$

From this equation, the standard deviations corresponding to each of the estimated variance components can be extracted and, depending on the relative magnitudes, inferences can be made regarding the 'goodness' of the estimated values. Another method, though beyond the scope of this paper, is the computation of the associated confidence intervals for each of the variance components.

In this study, a more empirical approach is followed whereby a number of different a-priori variance values for each height type are tested to see if they all yield the same solution. Interestingly, the results showed that the initial a-priori values for a single group of observations (e.g., N) do not have a noticeable effect on the remaining groups, even though theoretically co-dependence is present. It was also discovered that through proper VCE, the arbitrariness of certain pre-selected factors (such as

scaling $\mathbf{Q}_h$) can be alleviated. For instance, in the Swiss network, a factor of 10^2 was empirically determined as a scale for $\mathbf{Q}_h$. By re-scaling the original CV matrix with different a-priori factors, this independently-derived value was verified as an appropriate scale for $\mathbf{Q}_h$. This is an important realization as it ascribes some statistical reasoning towards an otherwise 'arbitrary' scaling practice and leads to a better understanding of the true measures of error.

5.2 Case Study II: Non-negative variances

One of the major pitfalls of the described VCE technique is that no provision has been made to ensure non-negative variance values (i.e. $\hat{\boldsymbol{\theta}} \in \Re_+^3$). In general, negative outcomes of variance components can be attributed to an insufficient number of observations compared to unknown parameters and/or an incorrect stochastic model. Thus, a negative variance outcome yields important information regarding the problem set up, information that is lost if the estimator is constrained to give only positive outcomes. Nonetheless, alternate algorithms exist which are constrained to provide positive variance factors. One such algorithm, known as iterative almost unbiased estimation (IAUE), can be implemented through the following formula [Rao and Kleffe, 1988]

$$\hat{\boldsymbol{\theta}}^{\mathrm{i}}_{new} = \frac{\hat{\boldsymbol{\theta}}^{\mathrm{i}}_{old}\mathbf{l}^{\mathrm{T}}\mathbf{R}\mathbf{Q}_{\mathrm{i}}\mathbf{R}^{\mathrm{T}}\mathbf{l}}{tr(\mathbf{R}\mathbf{Q}_{\mathrm{i}})} \qquad (13)$$

where $i = h, H, N$, $\hat{\boldsymbol{\theta}}^{\mathrm{i}}_{new}$ and $\hat{\boldsymbol{\theta}}^{\mathrm{i}}_{old}$ represent the current and previous iteratively-derived variance estimates respectively, and all other terms have been previously described. This algorithm was tested with the real data and proved to give almost identical results to those obtained using the I-MINQUE method. An added benefit of this method is that it is computationally simpler and converges approx-imately 50% faster than the other approach. Thus, in cases where computational efficiency is an issue, IAUE offers a viable alternative to I-MINQUE.

To determine the effect of data redundancy on the estimated variance components, using the iterated MINQUE approach, a test was performed whereby observations were eliminated (one-by-one) and the final values of the estimated components were noted. For the case of the Swiss test network, it was found that at least 49 observations are required for convergence and positive-valued final components, which corresponds to 44% of the available GPS-on-benchmarks in this region.

5.3 Case Study III: Effect of correlations

In practice, the availability of fully populated variance-covariance matrices for all groups of height data at the same GPS benchmarks is a luxury. To test the effect of correlations between observations of the same type on the estimated variance components, numerical experiments were conducted with fully populated and diagonal covariance matrices. The results are given in Table 2 (note that the estimated variance components are unitless scale factors). It is evident that due to the correlated nature of the observations, a diagonal covariance matrix is further from the 'true' CV matrix and therefore requires more iterations to obtain the final estimated values. Also, by neglecting the off-diagonal elements, we obtain overly optimistic CV matrices compared to the fully populated case. Results will vary depending on the degree of correlation, however it is clear that unrealistically 'good' results are obtained when correlations are ignored, as expected.

Table 2. Effect of correlations on estimated variances

Covariance Matrices	$\hat{\sigma}_h^2$	$\hat{\sigma}_H^2$	$\hat{\sigma}_N^2$	iterations
Full	2.83	5.06	1.02	99
Diagonal	0.71	3.63	1.07	152

5.3 Case Study IV: Parametric model type

As mentioned in the introduction, the optimal combination of h, H and N requires the incorporation of a parametric model to deal with the systematic errors inherent in the data and the fact that each height type refers to a different reference surface. The parametric model, or corrector surface

as it is commonly termed, can be incorporated as follows (Kotsakis and Sideris, 1999):

$$\mathbf{h} - \mathbf{H} - \mathbf{N} = \mathbf{A}\mathbf{x} + \mathbf{v} \qquad (14)$$

where the parametric term **Ax** describes the corrector surface. VCE procedures pre-suppose that *no biases* or *systematic effects* are present in the vector **v**. Any unmodelled effects may propagate into the estimated variances and give unreliable results. This case study was designed to determine the role of the parametric model type, if any, on the final estimated variance components. Five models were selected, namely the classic 4-parameter transformation model, as shown in the following formula

$$p_C(\varphi,\lambda) = x_o + x_1 \cos\varphi\cos\lambda + x_2 \cos\varphi\sin\lambda + x_4 \sin\varphi \qquad (15)$$

and four simple nested polynomial regression models from first to fourth order as given by

$$p_{nm}(\varphi,\lambda) = \sum_{m=0}^{M}\sum_{n=0}^{N} x_{nm}(\varphi-\varphi_o)^n(\lambda-\lambda_o)^m \qquad (16)$$

where φ,λ are the horizontal coordinates and φ_o,λ_o represent the midpoint of the network. The estimated variance components (unitless) corresponding to each parametric model type are shown in Table 3.

Table 3. Estimated variance components

Parametric Model Type	$\hat{\sigma}_h^2$	$\hat{\sigma}_H^2$	$\hat{\sigma}_N^2$
Eq. 15	2.82	5.06	1.01
Eq. 16, 1st order	2.83	4.80	1.01
2nd order	2.94	4.50	1.05
3rd order	3.08	3.96	0.97
4th order	divergence, negative estimates		

Predictably, a solution was not achievable in all cases leading to divergence and/or negative estimates. This may indicate that an inadequate model was used for the systematic effects resulting in 'residual' biases that corrupt the performance of the VCE method. Another possibility is the presence of numerical instabilities caused by over-parameterization, which occurred when a full 4th order model was used (Eq. 16, 4th order). In any case, these first results are revealing as they hint towards a means for identifying the effectiveness of the selected corrector surface. The latter comment is stated with prudence as further studies are currently underway to verify this deduction.

6 Conclusions and Future Work

The implementation of VCE in the optimal combined adjustment of heterogeneous height data, namely GPS, orthometric and geoid heights was presented. The I-MINQUE procedure was described and used to test the supplied CV matrices for each height type. Through various numerical case studies with real data, a number of key issues involved in most height-related applications were studied in detail. In the future, studies will be conducted in order to apply this approach for assessing the accuracy of global gravity field models (i.e. EGM96 and those obtained from GRACE/GOCE).

Acknowledgements

The authors would like to thank Urs Marti from the Swiss Federal Office of Topography for providing the GPS/levelling benchmark data in Switzerland.

References

Crocetto N, Gatti M and P Russo (2000) Simplified formulae for the BIQUE estimation of variance components in disjunctive observation groups. *Journal of Geodesy*, vol. 74: 447-457.

Grafarend E (1985) Variance covariance estimation, theoretical results and geodetic applications. *Statistics and Decisions*, Suppl. Issue 2: 407-441.

Kotsakis C and Sideris MG (1999) On the adjustment of combined GPS/levelling/geoid networks. *Journal of Geodesy*, vol. 73: 412-421.

Marti U (2002) Modelling of differences of height systems in Switzerland. *Proceedings of the 3rd Meeting of the International Gravity and Geoid Commission,* Tziavos (Ed.), Thessaloniki, Greece, August 2002: 378-388.

Rao CR (1971) Estimation of variance and covariance components MINQUE Theory. *Journal of Multivariate Analysis*, vol. 1: 257-275.

Rao CR and J Kleffe (1988) Estimation of Variance Components and Applications. North-Holland Series in Statistics and Probability, vol. 3, Amsterdam.

A minimum energy condition for the inverse gravimetric problem

Olivier Jamet
Laboratoire de Recherche en Géodésie, E.N.S.G.,
6 et 8 av. Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 Marne-La-Vallée CEDEX, France

Michel Diament
Laboratoire de Gravimétrie et de Géodynamique, UMR 7096, I.P.G.-P., 4 place Jussieu, 75005 Paris, France

Abstract. The problem of finding the density distribution of the Earth from gravity data is called the inverse gravimetric problem. It is well known that this problem has not a unique solution. A possible approach to force the uniqueness of the solution is to impose the solution to realize a minimum of energy, that is to be an equilibrium distribution.
Several authors have considered this possibility by setting the energy to the sole energy of the potential field. In this paper, we show that such approaches correspond to the maximization of a quadratic form in the set of the density distributions considered as an Hilbert space. As a consequence, they cannot deliver realistic results without a constraint that reduces the set of possible density distributions to a bounded domain of the considered Hilbert space. Moreover, under such constraints, the result will necessarily be located at the boundary of the searched domain.
In order to express a steady state condition leading to a freer solution, we propose to introduce a compression energy. We choose the state equation of Murnaghan that is known to apply reasonably in the upper mantle. We give an integral form of the compression energy and show that the resulting total energy admits a lower bound, which assess the existence of a solution. Finally, we derive a differential equation verified by that the absolute minima from the equilibrium equation under its local form.

Keywords. Inverse gravimetric problem, density distribution

1 Introduction

It is well known that finding the density distribution inside the Earth from its gravitation potential outside — that is solving the so called inverse gravimetric problem — is an ill-posed problem. The density distribution cannot be determined without additional conditions forcing the uniqueness of the solution. In this study, we focus on the approach that consists in considering that the matter inside the Earth is at a steady state, i.e. that it realizes a minimum of energy.

This approach was previously developed by Helmut Moritz (1990) who considered the energy of the Earth in its own gravitation field. This energy E_p, called the potential energy, can be expressed

$$E_p(\rho) = -\int_{\mathcal{V}} \rho(y) \left(\frac{1}{2} V_\rho(y) + \Phi(y) \right) dv_y \quad (1)$$

where $\mathcal{V}$ denotes the fixed volume of the Earth, $\rho(y)$ the value of the density at any point y of the volume $\mathcal{V}$, V_ρ and Φ respectively the gravitational and the centrifugal potential of the Earth, and dv_y the integration volume element at point y (in his book, H. Moritz does not consider the sign of this expression, but searches for extrema instead of minima).

It is obvious that minimizing the sole potential energy would not make much sense : without any other constraint, the matter would collapse toward the center of the Earth. H. Moritz proposed to solve for the extrema of the potential energy under a side condition saying that the volume enclosed by any surface of constant density remains unchanged during the variation to be performed. Formulated this way, the inverse gravimetric problem could yield realistic density distributions. Though, as F. Sansò remarked recently (Sansò 2001), this problem remains unsolved.

In this paper, we consider the simpler case of a non rotating Earth ($\Phi \equiv 0$) and discuss this problem from a mathematical and a physical point of view. Firstly, we give a geometric interpretation to the minimization of the potential energy within the Hilbert space of all the density distributions defined in the volume $\mathcal{V}$ (section 2). This interpretation questions the physical meaning of the resulting minima. Secondly, we propose to give the inverse gravimetric problem a new formulation based on a physical model of the equilibrium, by introducing the lithostatic pressure forces that counterbalance the gravitational forces (section 3). Theses pressure forces are modeled though the introduction of an equation of state. We choose the

equation of Murnaghan (Poirier 2000) that applies in the upper mantle and show that the minimization problem admits solutions. Finally, in the section 4, we show that the equation of state of Murnaghan leads to a rather simple differential equation on the density distribution.

2 A geometric interpretation

In the following, we consider the set $\mathcal{E}$ of real functions ρ defined in the volume $\mathcal{V}$. For the convenience of the study, we will consider theses functions as two times differentiable in the interior of $\mathcal{V}$. This property is not necessary for the discussion of the present section. It will only be useful in the sections 3 (where we need the first derivative) and 4 (where we need the second derivative). We also consider that theses function are square integrable over $\mathcal{V}$, which is of course the case, since they represent a physical source of finite energy.

Obviously, associated with its natural scalar product $\langle \rho_1, \rho_2 \rangle = \int_{\mathcal{V}} \rho_1 \rho_2$, $\mathcal{E}$ is a Hilbert space. Solving Moritz's inverse gravimetric problem for a non rotating Earth consists then in finding $\rho \in \mathcal{E}$ that verifies

$$\forall x \notin \mathcal{V},\ V_\rho(x) \equiv k \int_{\mathcal{V}} \frac{\rho(y)}{\|y-x\|} dv_y = u_{ext}(x) \tag{2}$$

where u_{ext} is the given gravitational potential outside of the Earth and k the universal gravitation constant, and that minimize the potential energy

$$E_p = -\frac{1}{2} \int_{\mathcal{V}} \rho(y)\, V_\rho(y)\, dv_y \tag{3}$$

$$= -\frac{k}{2} \int_{\mathcal{V}} \int_{\mathcal{V}} \frac{\rho(x)\,\rho(y)}{\|y-x\|} dv_x dv_y \tag{4}$$

2.1 Afine subspace of the admissible densities

Let us call $\mathcal{D}_{u_{ext}}$ the subset of $\mathcal{E}$ of the density distributions that verify the constraint equation 2, and consider two arbitrary members ρ_1 and ρ_2 of this subset. We will call such density functions admissible densities. Then,

$$\forall x \notin \mathcal{V},\ V_{\rho_2}(x) - V_{\rho_1}(x) = V_{\rho_2-\rho_1}(x) = 0$$

As a consequence, if we call $\mathcal{D}_0$ the subset of the density distributions that verify

$$\forall \rho \in \mathcal{D}_0,\ \forall x \notin \mathcal{V},\ V_\rho(x) = 0$$

for any density $\rho_1 \in \mathcal{D}_{u_{ext}}$, we can write

$$\forall \rho_2 \in \mathcal{D}_{u_{ext}},\ \exists \rho \in \mathcal{D}_0,\ \rho_2 = \rho_1 + \rho$$

or, in other words

$$\mathcal{D}_{u_{ext}} = \rho_1 + \mathcal{D}_0 \tag{5}$$

One can verify easily that $\mathcal{D}_0$ is a Hilbert subspace of $\mathcal{E}$. Hence $\mathcal{D}_{u_{ext}}$ is an afine subspace of $\mathcal{E}$.

2.2 Properties of the potential energy

Let us call $L_p(\rho_1, \rho_2)$ the integral

$$L_p(\rho_1, \rho_2) = k \int_{\mathcal{V}} \int_{\mathcal{V}} \frac{\rho_1(x)\,\rho_2(y)}{\|y-x\|} dv_x dv_y$$

Physically, this integral represents the opposite of the mutual potential energy of ρ_1 and ρ_2, that is the potential energy of the density distribution ρ_1 (resp. ρ_2) in the gravitational field of ρ_2 (resp. ρ_1). Of course, this physical interpretation remains abstract, since here the two density distributions occupy the same volume, which does not make sense in reality.

Obviously, L_p defines a symmetric bilinear form of the Hilbert space $\mathcal{E}$. Consequently, the integral Q_p defined as

$$Q_p(\rho) \equiv L_p(\rho, \rho) = -2E_p(\rho) \tag{6}$$

is the quadratic form of $\mathcal{E}$ associated to the bilinear form L_p. This quadratic form is positive definite. The proof of this latter property is exposed in the appendix. Hence, L_p defines a scalar product of $\mathcal{E}$, and $\sqrt{Q_p}$ a distance.

2.3 Alternative expression of $\mathcal{D}_{u_{ext}}$

Since Q_p is positive definite, there is a unique density distribution in $\mathcal{D}_{u_{ext}}$ for which Q_p is minimal. This density distribution is orthogonal to the leading Hilbert space $\mathcal{D}_0$ of $\mathcal{D}_{u_{ext}}$, in the sense of the scalar product L_p. Let us call ρ_M this function. One can then write

$$\mathcal{D}_{u_{ext}} = \rho_M \oplus_{L_p} \mathcal{D}_0 \tag{7}$$

where $\oplus_{L_p}$ designate the orthogonal sum according to the scalar product L_p.

2.4 Interpretation of Moritz's approach

The inverse gravimetric problem that consists in minimizing the potential energy of equation 3 under the constraint 2 can hence be interpreted as the maximization of a positive definite quadratic form — or as the maximization of the associated distance — in an afine subspace of $\mathcal{E}$.

As said in introduction from physical reasoning, such a problem cannot have a meaningful solution

since an afine space is unbounded. Consequently, a « side constraint » has to be specified. However, one can now understand that, whatever be the side constraint, it will give a meaning to the problem if and only if it defines a bounded subset of $\mathcal{D}_{u_{ext}}$. And if it does, maximizing a distance within this bounded set will necessarily locate the solution at its boundary. In that sense, the constraints that can be introduced to solve the inverse gravimetric problem will always highly influence the solution, which questions its physical meaning.

We must note that, in the case of a rotating Earth, the opposite of the potential energy will not be definite and this conclusion drops. However, considering the respective order of magnitude of the gravitational energy and of the centrifugal energy, one can expect that event with rotation, the solution will be located very near from the boundary of the domain defined by the side constraint.

This is why we have tried to reformulate the inverse gravimetric problem by taking into account other forces, in order to avoid the need for constraints bounding the solutions to a search domain of finite extent.

2.5 Additional considerations

It is remarkable to note that the present geometric reasoning gives another demonstration of a previously established result. When considering the ellipsoid in its normal potential field, the density distribution which maximizes the potential energy is the equivalent surface layer generating the potential u_{ext}. This property was commented by F. Sansò (2001), and was established by solving the Euler equations corresponding to the stationarity of E_p with 2 as condition. Here, we obtain directly this result by remarking

- that the potential u_{ext} can be generated by a single layer located at the boundary of $\mathcal{V}$; this is a direct consequence of Green's third identity (see (Blakely 1996, p. 25)) applied to the case when the boundary of the domain where the potential field is harmonic is equipotential ; moreover, this single layer is uniquely defined by the sole knowledge of u_{ext} ;
- that any single layer located at the boundary of $\mathcal{V}$ is orthogonal (according to the scalar product L_p) to $\mathcal{D}_0$; indeed, if ρ_s is such a single layer

$$\forall \rho_0 \in \mathcal{D}_0, L_p(\rho_s, \rho_0) = \int_{\mathcal{V}} \rho_s V_{\rho_0} = 0$$

since ρ_s is null inside $\mathcal{V}$, and V_{ρ_0} is null on the boundary of $\mathcal{V}$.

3 Modeling the lithostatic pressure

The above considerations lead us to search for a better formulation of the inverse gravimetric problem. Since in the reality the gravitational collapse is counterbalanced by the lithostatic pressure, it was natural to try to introduce this term in the expression of the energy to be minimized. It is the purpose of the present section.

3.1 The inverse gravimetric problem in differential form

Let us call $p(x)$ the pressure field inside the volume $\mathcal{V}$. In the absence of any other force, the steady state condition at any point of the volume can be written

$$\forall x \in \mathcal{V},\ \rho(x)\nabla V_\rho(x) - \nabla p(x) = 0 \qquad (8)$$

This condition is sufficient to define the equilibrium under the hypothesis of a relation between the local values of the density and pressure, that is namely of an equation of state. In this study, we chose the equation of state of Murnaghan for a reason of simplicity. This equation depends on only three parameters — the values, at the atmospheric pressure, of the density ρ_0, the bulk modulus K_0, and the first derivative K_0' of the bulk modulus considered as a function of the pressure — and gives a simple expression of the pressure as a function of the density (Poirier 2000) :

$$p(x) = \frac{K_0}{K_0'}\left(\left(\frac{\rho(x)}{\rho_0}\right)^{K_0'} - 1\right) \qquad (9)$$

Besides, this equation applies reasonably to the upper mantle (Poirier 2000), and thus gives a sound model of the nearest sources (at the scale of the Earth, we will neglect the crust). Of course, choosing a single equation of state for the whole Earth is a very coarse simplification. A realistic model should at least handle separately the four main states of the matter inside the Earth (solid core, fluid core, lower and the upper mantle).

If we accept this simplification, the inverse gravimetric problem consists then in finding the density distribution ρ that satisfies the two local equations 8 and 9 under the constraint 2. That is what we call the differential form of the inverse gravimetric problem.

3.2 The inverse gravimetric problem in integral form

The steady state condition 8 corresponds to a minimum of the total energy defined as the sum of the compression energy and the potential energy. In this section we give an expression of the compression energy, under the hypothesis of the equation of state of Murnaghan and express the inverse gravimetric problem as a minimization problem.

3.2.1 Compression energy

The compression energy can easily be computed, for a unit mass, as the work of the pressure forces during the compression from its volume v_0 under the atmospheric pressure to the volume v_1. We suppose this compression adiabatic. This hypothesis is reasonable if we suppose that the matter collapses sufficiently fast for getting to the steady state without thermal exchanges. This hypothesis is also necessary since we do not handle the thermal energy in the model. The work w_ρ per unit of mass of the pressure forces during this volume change can be calculated as

$$w_\rho(v_1) = -\int_{v_0}^{v_1} p dv$$

By substituting $\rho = \frac{1}{v}$ in this expression and using the equation of state 9, we get

$$\begin{aligned} w_\rho(\rho_1) &= \int_{\rho_0}^{\rho_1} \frac{p}{\rho^2} d\rho \\ &= \frac{K_0}{K_0'}\left(\frac{\rho_1^{K_0'-1}}{(K_0'-1)\rho_0^{K_0'}} + \frac{1}{\rho_1} - \frac{K_0'}{(K_0'-1)\rho_0}\right) \end{aligned}$$

The work $w_v(\rho_1)$ per unit of volume during this compression can then be calculated as

$$w_v(\rho_1) = \rho_1 w_\rho(\rho_1)$$

and the total work W_t on the entire volume $\mathcal{V}$ as

$$\begin{aligned} W_t &= \int_{\mathcal{V}} w_v(\rho(x))\, dv_x \\ &= \frac{K_0}{K_0'(K_0'-1)\rho_0^{K_0'}} \int_{\mathcal{V}} \rho(x)^{K_0'} .dv_x \\ &\quad + \frac{K_0}{K_0'} v_{\mathcal{V}} - \frac{K_0}{(K_0'-1)\rho_0} m_{\mathcal{V}} \end{aligned}$$

where $v_{\mathcal{V}}$ is the volume of $\mathcal{V}$ and $m_{\mathcal{V}}$ is the total mass of $\mathcal{V}$.

Since the volume $\mathcal{V}$ is fixed, and since under the constraint given by the equation 2 the total mass of $\mathcal{V}$ is also fixed, we can drop the last two terms in the expression of the total compression work (they will not interfere with the minimization), and define the compression energy E_c as

$$E_c = \frac{K_0}{K_0'(K_0'-1)\rho_0^{K_0'}} \int_{\mathcal{V}} |\rho(x)|^{K_0'} .dv_x$$

where we take the absolute value of the density inside the integral in order to give sense to this formula inclusively for the negative densities (that are considered as soon as the problem is expressed within an Hilbert space).

3.2.2 A new minimization problem

The inverse gravimetric problem can thus be expressed as the search for the density distribution that minimizes the total energy E_t defined as

$$\begin{aligned} E_t &= \frac{K_0}{K_0'(K_0'-1)\rho_0^{K_0'}} \int_{\mathcal{V}} |\rho(x)|^{K_0'} .dv_x \\ &\quad -\frac{1}{2}\int_{\mathcal{V}} \rho(x) V_\rho(x)\, dv_x \end{aligned} \tag{10}$$

under the constraint given by equation 2.

3.3 Existence of solutions

The integral form gives us an immediate proof for the existence of solutions. Indeed, in the appendix, we show that the quadratic form Q_p admits an upper bound, function of the traditional square norm of the density distribution

$$\exists a > 0, \int_{\mathcal{V}} \rho V_\rho \le a \int_{\mathcal{V}} \rho^2 \tag{11}$$

Consequently, the total energy E_t verifies

$$\begin{aligned} E_t &> \frac{K_0}{K_0'(K_0'-1)\rho_0^{K_0'}} \int_{\mathcal{V}} |\rho|^{K_0'} .dv_x \\ &\quad -a\int_{\mathcal{V}} \rho^2 .dv_x \\ &= \frac{K_0}{K_0'(K_0'-1)\rho_0^{K_0'}} \int_{\mathcal{V}} \left(|\rho|^{K_0'} - b\rho^2\right) dv_x \end{aligned}$$

where b is a positive constant. This latter expression admits a minimum as soon as $K_0' > 2$. This condition is only a sufficient condition. Firstly, we did not examine the case $K_0' = 2$ (for which the lower bound could admit a minimum). Secondly, Q_p may admit smaller upper bounds that the one we used.

Nevertheless, this condition is sufficient for our problem. Geophysical evidence show that the value of K_0' for all the rocks of the upper mantle is very close from 4 (Poirier 2000). Hence, the total energy does admit an absolute minimum.

4 A local steady state condition

In the previous sections we proposed two equivalent formulations of an inverse gravimetric problem taking into account the lithostatic pressure forces. While the integral form is convenient for assessing the existence of at least one solution, it is not obvious that it be the better way to investigate its possible uniqueness. The total energy is the difference of an integral of the power 4 (approximately) of the density values and a quadratic form. It is well known that even in dimension 0 (that is considering polynomials instead of integrals), such an expression may admit several absolute minima. The possible uniqueness would come from some asymmetry in the problem, that may be difficult to establish in a space of infinite dimension.

Thus, it might be more convenient to look for the solutions of the problem under its differential form. If we admit that the density distributions are twice differentiable, the equations 8 and 9 lead to a rather simple steady state equation on the sole density. Indeed, injecting the expression of the pressure from equation 9 into the equation 8 leads to

$$\begin{aligned}\rho\nabla V_\rho &= \nabla\left(\frac{K_0}{K_0'}\left(\left(\frac{\rho}{\rho_0}\right)^{K_0'}-1\right)\right)\\ &= \frac{K_0}{\rho_0^{K_0'}}\rho^{K_0'-1}\nabla\rho\end{aligned}$$

which becomes, in the part of the volume where the density is non null (the Murnaghan equation is anyhow valid only there)

$$\nabla V_\rho = \frac{K_0}{\rho_0^{K_0'}}\rho^{K_0'-2}\nabla\rho$$

By taking the divergence of this latter expression, we get, using the Poisson relation,

$$\frac{\rho_0^{K_0'}}{4\pi k K_0}\rho + \nabla.\left(\rho^{K_0'-2}\nabla\rho\right) = 0$$

which can be developed into (where the density is non null)

$$\rho^{K_0'-3}\nabla^2\rho + (K_0'-2)\,\rho^{K_0'-4}\,(\nabla\rho)^2 = -\frac{\rho_0^{K_0'}}{4\pi k K_0} \tag{12}$$

In the case $K_0' = 4$, which is a reasonable value for the upper mantle, this equation becomes

$$\rho\nabla^2\rho + 2\,(\nabla\rho)^2 = -\frac{\rho_0^4}{4\pi k K_0} \tag{13}$$

Such an equation has to be solved under the constraint given by equation 2, that is for $\rho \in \mathcal{D}_{u_{ext}}$. Using the direct decomposition expressed in equation 7, equation 13 becomes

$$(\rho+\rho_M)\,\nabla^2\,(\rho+\rho_M) + 2\,(\nabla\,(\rho+\rho_M))^2 = -c$$

where c is a positive constant, ρ_M is a given density distribution, and $\rho \in \mathcal{D}_0$.

5 Conclusion

In this paper, we discussed the inverse gravimetric problem — in the case of a non rotating Earth — as the search for a density distribution satisfying a given potential outside the volume $\mathcal{V}$ of the sources within the Hilbert space of the real functions defined in $\mathcal{V}$. We showed that the potential energy is proportionnal to a quadratic form of that space, and that minimizing the potential energy is equivalent to maximizing a distance, which cannot make sense without an additional constraint such as the one proposed by H. Moritz (1990). Moreover, when the constrained problem admits a solution, the latter is necessarily located at the boundary of the domain defined by the given constraint in the considered Hilbert space.

We then proposed an alternative consisting in taking into account the lithostatic pressure. We proposed to consider the equation of state of Murnaghan as applying in the whole volume and gave the expression of the total energy defined as the sum of the compression energy and the potential energy under this assumption. We demonstrated the existence of an absolute minimum. Finally, we showed that the problem could also be tackled though the resolution of a local differential equation on the density distribution.

We are far from a proof of the uniqueness of the solution under such assumptions, and still further from a numerical resolution of the problem in the case of the real Earth. However, this alternative to the classical formulation of the inverse gravimetric problem in geodesy opens new perspectives. The computation of the steady state of the ellipsoid could profit by former works in the community of the astrophysicists who have tackled for long a similar problem for the stars in the computation of their gravitational collapse. As for the real Earth, the problem remain rather different since the boundary constraint is much more complex. We hope that the geometrical interpretation given in this paper will be of some help for simplifying its resolution.

References

Blakely R. (1996) *Potential theory in gravity and magnetic applications,* Cambridge University Press.

Moritz H. (1990) *The figure of the Earth, Theoretical geodesy and the Earth's interior*, Wichmann, Karlsruhe.

Poirier J.-P. (2000) *Introduction to the physics of the Earth's interior*, Cambridge University Press.

Sansò F. (2001) The greening of geodetic theory, *Proc. Scientific Assembly of IAG, Budapest, 2-7 Sept.*, International Association of Geodesy Symposia, vol. 125, Springer-Verlag, Berlin, pp. 595-601

Appendix

In this appendix, we give the proof that the quadratic form Q_p introduced in section 2 is positive definite. Additionally, we show that Q_p admits an upper bound expressed in function of $\int_{\mathcal{V}} \rho^2$.

Proof for positive definiteness. With the variable change $h = y - x$, Q_p can be expressed

$$Q_p(\rho) = k \int_{x \in \mathcal{V}} \int_{x+h \in \mathcal{V}} \frac{\rho(x)\,\rho(x+h)}{\|h\|} dv_x dv_h$$

As ρ is null outside $\mathcal{V}$ and $\mathcal{V}$ is bounded, if we define $\mathcal{W}$ as the spherical volume of radius A for which

$$\forall x, y \in \mathcal{V},\ y - x \in \mathcal{W}$$

$\mathcal{W}$ is also bounded and we have

$$\begin{aligned} Q_p(\rho) &= k \int_{x \in \mathcal{W}} \int_{h \in \mathcal{W}} \frac{\rho(x)\,\rho(x+h)}{\|h\|} dv_x dv_h \\ &= k \int_h dv_h \frac{1}{\|h\|} \int_x dv_x \rho(x)\,\rho(x+h) \\ &= k \int dv_h \varphi(h)\,\gamma(h) \end{aligned}$$

where the last integral is unbounded and where we define the function φ as

$$\begin{cases} h \in \mathcal{W} : \varphi(h) = \frac{1}{\|h\|} \\ h \notin \mathcal{W} : \varphi(h) = 0 \end{cases}$$

and the function γ as the 3D autocorrelation function of the density distribution ρ

$$\gamma(h) = \int_{x \in \mathcal{W}} dv_x \rho(x)\,\rho(x+h)$$

In consequence of the hypothesis taken on the density distributions (section 2), the function γ is square integrable. As for the function φ, the integral of φ^2 converges near the origin since the volume element dv_h behaves as h^2. Besides, since φ is finite everywhere else, and non null on a finite domain, it is also square integrable. Thus both function admit a finite 3D Fourier transform that will be noted Φ and Γ respectively. Γ being the Fourier transform of the autocorrelation function of ρ, if we note R the 3D Fourier transform of ρ, we have $\Gamma(u) = R(u)\,R^*(u)$ (with R^* beeing the conjugate of R), and thus

$$\begin{aligned} Q_p(\rho) &= k \int dv_h \varphi(h)\,\gamma(h) = k\,(\Phi * \Gamma)(0) \\ &= k \int dv_u R(u)\,R^*(u)\,\Phi(u) \end{aligned} \qquad (14)$$

(where $\Phi * \Gamma$ denotes their convolution product).

The announced property comes then from the expression of $\Phi(u)$. Calculated in spherical coordinates, with the notation $h = (r, \theta, \lambda)$ and $u = (\eta, \Theta, \Lambda)$, we have

$$\Phi(\eta, \Theta, \Lambda) = \int r \sin(\theta)\, \mathrm{e}^{-2i\pi(u.h)} dr.d\theta.d\lambda$$

which gives by a change of the angular origin

$$\begin{aligned} \Phi(\eta, \Theta, \Lambda) &= \int r \sin(\theta)\, \mathrm{e}^{-2i\pi\eta r \cos(\theta)} dr d\theta d\lambda \\ &= \frac{1}{\pi\eta^2}(1 - \cos(2\pi A\eta)) \end{aligned}$$

Thus Φ is positive and non null everywhere except on a countable number of spheres $\eta = \frac{n}{A}$ (n integer). Hence, on one hand, Q_p is necessarily positive, and on the other hand :

$$Q_p(\rho) = 0 \ \Rightarrow \ \begin{cases} \forall u = (\eta, \Theta, \Lambda), \\ \eta \neq \frac{n}{A} \Rightarrow R(u) = 0 \end{cases}$$

which is sufficient to ensure that $R(u) \equiv 0$, since this property applies for any A greater than a minimum value. Hence, $Q_p(\rho) = 0$ implies that ρ is null.

Upper bound of Q_p. Moreover, from the equation 14, and from the fact that Φ is bounded, we have

$$\exists a > 0,\ \forall u, 0 \leq \Phi(u) \leq a$$

and consequently (Parseval theorem)

$$\begin{aligned} Q_p(\rho) &\leq a.k \int dv_u R(u)\,R^*(u) \\ &= a.k \int dv_x \rho^2(x) \end{aligned}$$

Successive Approximations in the Solution of a Weakly Formulated Geodetic Boundary-Value Problem

P. Holota
Research Institute of Geodesy, Topography and Cartography, 250 66 Zdiby 98, Praha-vychod, Czech Republic
e-mail: holota@pecny.asu.cas.cz, Tel.: +420 323 649235, Fax: +420 2 84890056

Abstract. For many purposes a sphere or an ellipsoid of revolution approximate the figure of the Earth fairly well. However, the Earth's topography causes effects that continuously stimulate a considerable research effort. In contrast to the concept of the so-called shrinking parameter in the classical treatment of the effect, successive approximations are applied within a weak formulation of the respective boundary-value problem in this paper. In this connection functional-analytic estimates are investigated so as to clarify the convergence of the iteration process. In particular the ellipticity of a bilinear form associated with the problem under consideration is discussed for the solution domain of an ellipsoidal boundary. The apparatus of ellipsoidal harmonics is substantially used for this purpose. Quantitative estimates for the respective parameters were derived, when the potential is assumed to follow a finite degree model.

Keywords. Earth's gravity field, geodetic boundary-value problems, variational methods, functional-analytic estimates

1 Introduction

The purpose of this paper is to discuss an iteration approach to the solution of a geodetic boundary-value problem in the exterior of the Earth, i.e. in an unbounded domain Ω. In this way we attempt to treat the effect of the irregularity of the boundary caused by its topography. Our aim is to formulate the process of successive approximations within the concept of the weak solution and variational principles, see Necas (1967) and Rektorys (1974).

Let us start with a quadratic functional

$$\Phi(u) = A(u,u) - 2\int_{\partial\Omega} uf\, dS \qquad (1)$$

defined for all functions u from Sobolev's space $W_2^{(1)}(\Omega)$ endowed with inner product

$$(u,v)_{W_2^{(1)}(\Omega)} \equiv \int_{\Omega} \frac{uv}{r^2}\, dx + \sum_{i=1}^{3} \int_{\Omega} \frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_i}\, dx \qquad (2)$$

where x_i, $i = 1,2,3$, are rectangular Cartesian coordinates in Euclidean three-dimensional space $\boldsymbol{R}^3$, $r = |\boldsymbol{x}| = (\Sigma_{i=1}^{3} x_i^2)^{1/2}$ and $dx = dx_1 dx_2 dx_3$ means the volume element;

$$A(u,v) = \int_{\Omega} \langle \boldsymbol{grad}\, u, \boldsymbol{grad}\, v \rangle\, dx \qquad (3)$$

is a bilinear form defined on the Cartesian product $W_2^{(1)}(\Omega) \times W_2^{(1)}(\Omega)$, $\langle .,. \rangle$ means the inner product of two vectors and f belongs to $L_2(\partial\Omega)$, i.e. to the space of square-integrable functions on $\partial\Omega$, where $\partial\Omega$ is the boundary of Ω. As to the regularity, we suppose that $\partial\Omega$ is Lipschitz' boundary.

The functional Φ attains its minimum in $W_2^{(1)}(\Omega)$, which results from the theory of nonlinear functionals and the solution of an abstract variational problem, as e.g. in Necas and Hlavacek (1981). In our particular case the respective reasoning may be found in Holota (2000).

Following now the concepts of calculus of variations and assuming that Φ has its local minimum at a point $u \in W_2^{(1)}(\Omega)$, we necessarily arrive at

$$A(u,v) = \int_{\partial\Omega} vf\, dS \qquad (4)$$

valid for all $v \in W_2^{(1)}(\Omega)$, see Holota (2000). This integral identity represents Euler's necessary condition for Φ to have a minimum at the point u. It has also a classical interpretation under some regularity assumptions, which shows that u has to be a solution of Neumann's boundary-value problem, i.e.,

$$\Delta u = 0 \text{ in } \Omega \text{ and } \frac{\partial u}{\partial n} = -f \text{ on } \partial\Omega \qquad (5)$$

where Δ means Laplace's operator and $\partial/\partial n$ denotes the derivative in the direction of the unit normal $\boldsymbol{n}$ of $\partial\Omega$, see Holota (2000).

For geodetic applications, however, we have to consider a more general case. We have to take $A(u,v) = A'(u,v) - A''(u,v)$, where

$$A'(u,v) = \int_{\Omega} \langle \mathbf{grad}\, u, \mathbf{grad}\, v \rangle \, dx \tag{6}$$

$$A''(u,v) = \int_{\Omega} \langle \mathbf{grad}\, v, \boldsymbol{a} \times \mathbf{grad}\, u \rangle \, dx + \int_{\Omega} v \langle \mathbf{curl}\, \boldsymbol{a}, \mathbf{grad}\, u \rangle \, dx \tag{7}$$

and $\boldsymbol{a} = (a_1, a_2, a_3)$ is a vector field such that the components a_i and also $|\boldsymbol{x}|(\mathbf{curl}\, \boldsymbol{a})_i$, $i = 1,2,3$, are Lebesgue measurable functions defined and bounded almost everywhere on Ω, see Holota (1997). The integral identity

$$A(T,v) = \int_{\partial\Omega} vf \, dS \tag{8}$$

valid for all $v \in W_2^{(1)}(\Omega)$ then defines a function $T \in W_2^{(1)}(\Omega)$ which is a generalized solution of the following classical problem

$$\Delta T = 0 \quad \text{in} \quad \Omega \tag{9}$$

$$\langle \boldsymbol{s}, \mathbf{grad}\, T \rangle = f \quad \text{on} \quad \partial\Omega \tag{10}$$

provided that on $\partial\Omega$ the vector $\boldsymbol{\sigma} = \boldsymbol{s}/\langle \boldsymbol{s}, \boldsymbol{n} \rangle$ and the field $\boldsymbol{a}$ are coupled so that $\boldsymbol{\sigma} = \boldsymbol{n} + \boldsymbol{a} \times \boldsymbol{n}$.

This problem has an important geodetic interpretation. Indeed, let δg be the gravity disturbance and U the standard potential of the Earth, so that $\gamma = |\mathbf{grad}\, U|$ is the normal gravity. We assume that δg is corrected for the gravitational interaction with the Moon, the Sun and the planets, for the precession and nutation of the Earth, for the atmospheric attraction and so on. Putting now $f = -\delta g$ and $\boldsymbol{s} = -(1/\gamma)\, \mathbf{grad}\, U$, we can see that Eqs. (9) and (10) define the linearized gravimetric fixed boundary-value problem and we know that its solution T is the disturbing potential, cf. Holota (1997), which focuses on the weak theory of the problem. This theory was also discussed in Rozanov and Sansò (2003).

It has been shown in Holota (2000) that the problem may be solved by means of successive approximations. Within this approach we consider a sequence $[u_m]_{m=0}^{\infty}$ of functions defined by

$$A'(u_{m+1}, v) = \int_{\partial\Omega} vf \, dS + A''(u_m, v) \tag{11}$$

which is assumed to hold for all $v \in W_2^{(1)}(\Omega)$ and $m = 0,1,\ldots,\infty$. It has been then shown that $[u_m]_{m=0}^{\infty}$ is a Cauchy (fundamental) sequence in $W_2^{(1)}(\Omega)$, so that it converges to a function $u \in W_2^{(1)}(\Omega)$ in the topology of that space. In addition the iterations may be also expressed by the relation

$$A'(u_{m+1}, v) = \int_{\partial\Omega} v f_m \, dS \tag{12}$$

assumed to hold for all $v \in W_2^{(1)}(\Omega)$, while $f_m = f - (\partial u_m / \partial t)\tan(\boldsymbol{s}, \boldsymbol{n})$ and $\partial / \partial t$ denotes the derivative in the direction of $\boldsymbol{t} = (\boldsymbol{\sigma} - \boldsymbol{n})/|\boldsymbol{\sigma} - \boldsymbol{n}|$. Note that $\boldsymbol{t}$ is tangential to $\partial\Omega$ (and exists almost everywhere on a Lipschitz' boundary). Details on the geometrical considerations in the computation of the correction term $(\partial u_m / \partial t)\tan(\boldsymbol{s}, \boldsymbol{n})$ and the use of components of the deflections of the vertical have been also discussed in Holota (1997, 2000).

Eq. (12) has the same structure as Euler's necessary condition for Φ to have a minimum at the point u_{m+1}. Thus every iteration step may be interpreted as a minimization problem, which represents a generalization of Neumann's problem, as we already saw.

2 Linear system

It is worth mentioning that u_{m+1} is defined as an element of a function space which is even too large. Note that in our case this does not imply a serious complication. It is enough to consider just a space $H_2^{(1)}(\Omega)$ of those functions from $W_2^{(1)}(\Omega)$ which are harmonic in Ω and to reformulate our problem, i.e. to look for $u_{m+1} \in H_2^{(1)}(\Omega)$ such that Eq. (12) holds for all $v \in H_2^{(1)}(\Omega)$. Moreover, we can approximate u_{m+1} by means of

$$u^{(n,m)} = \sum_{j=1}^{n} c_j^{(n,m)} v_j \tag{13}$$

where v_j are members of a function basis of $H_2^{(1)}(\Omega)$ and determine the coefficients $c_j^{(n,m)}$ as a solution of Galerkin's system

$$\sum_{j=1}^{n} c_j^{(n,m)} A'(v_j, v_k) = \int_{\partial\Omega} v_k f_m \, dS, \quad k = 1,\ldots,n \tag{14}$$

connected with integral identity (12).

Nevertheless, in practice a slight modification of the bilinear form $A'(u,v)$ is useful. In particular put

$$A^*(u,v) = \int_{\Omega^*} \langle \mathbf{grad}\, u, \mathbf{grad}\, v \rangle \, dx \tag{15}$$

for $u, v \in H_2^{(1)}(\Omega^*)$, where (in contrast to Ω) the domain Ω^* has a "simpler boundary". As v_j, however, we have to take members of a function basis of $H_2^{(1)}(\Omega^*)$. In $H_2^{(1)}(\Omega^*)$ they generate a sequence of finite dimensional subspaces

$$H_n(\Omega^*) = \text{span}\{v_i,\ i = 1,\ldots,n\},\ n = 1,2,\ldots \quad (16)$$

In the sequel we will suppose that $\Omega \subseteq \Omega^*$, which fairly well corresponds to the practical situation. Moreover, using Runge's property of Laplace's equation, see e.g. Bers et al. (1964), Krarup (1969) or Moritz (1980), we can take v_j, restricted to Ω, as members of a function basis in $H_2^{(1)}(\Omega)$ too. Under this restriction we put by analogy

$$H_n(\Omega) = \text{span}\{v_i,\ i = 1,\ldots,n\},\ n = 1,2,\ldots \quad (17)$$

and it is clear that $H_n(\Omega) = H_n(\Omega^*)\big|_{\Omega}$. Thus for all $v \in H_n(\Omega^*)$

$$A^*(u^{(n,m)}, v) = \int_{\partial\Omega} v f_m \, dS + F(v) \quad (18)$$

where

$$F(v) = A^*(u^{(n,m)}, v) - A'(u^{(n,m)}, v) = \int_D \langle \boldsymbol{grad}\, u^{(n,m)}, \boldsymbol{grad}\, v \rangle \, dx \quad (19)$$

and $D = \Omega^* - \Omega$. Obviously, Eqs. (18) and (19) imply a continuation of $u^{(n,m)}$, but we can take Eq. (18) as an alternative expression of the system given by Eqs. (14).

3 Successive approximations

Our aim is now to solve the integral identity (18) by means of successive approximations. We will examine the sequence of functions

$$u_0^{(n,m)},\ u_1^{(n,m)},\ \cdots,\ u_k^{(n,m)},\ u_{k+1}^{(n,m)},\ \cdots \quad (20)$$

defined by

$$A^*(u_{k+1}^{(n,m)}, v) = \int_{\partial\Omega} v f_m \, dS + F_k(v) \quad (21)$$

valid for all $v \in H_n(\Omega^*)$, where

$$F_k(v) = \int_D \langle \boldsymbol{grad}\, u_k^{(n,m)}, \boldsymbol{grad}\, v \rangle \, dx \quad (22)$$

The open question is whether the sequence $[u_k^{(n,m)}]_{k=0}^{\infty}$ has a limit. For this purpose we first deduce from Eq. (21) that

$$A^*(u_{k+2}^{(n,m)} - u_{k+1}^{(n,m)}, v) = \bar{F}(v) \quad (23)$$

where

$$\bar{F}(v) = \int_D \langle \boldsymbol{grad}\,(u_{k+1}^{(n,m)} - u_k^{(n,m)}), \boldsymbol{grad}\, v \rangle \, dx \quad (24)$$

holds for all $v \in H_n(\Omega^*)$. It is clear that for $u_{k+1}^{(n,m)} - u_k^{(n,m)}$ fixed, $\bar{F}(v)$ is a bounded linear functional of the variable v. Indeed, using Hölder's inequality, we easily deduce that

$$|\bar{F}(v)| \le M \sqrt{\int_D |\boldsymbol{grad}\, v|^2 \, dx} \le M \, \| v \|_{H_2^{(1)}(\Omega^*)} \quad (25)$$

where

$$M = \sqrt{\int_D |\boldsymbol{grad}\,(u_{k+1}^{(n,m)} - u_k^{(n,m)})|^2 \, dx} \quad (26)$$

Thus the norm $\| \bar{F} \|$ of $\bar{F}$ can be estimated from above by M, i.e. $\| \bar{F} \| \le M$. However, for subsequent estimates some finer considerations are needed.

It is obvious that for a function $v \in H_2^{(1)}(\Omega^*)$ which is not identically zero

$$\int_D |\boldsymbol{grad}\, v|^2 \, dx < \int_{\Omega^*} |\boldsymbol{grad}\, v|^2 \, dx \quad (27)$$

in view of the analyticity of v. But more than that, for functions $\bar{v} \in H_2^{(1)}(\Omega^*)$ of equally limited spectrum (band limited spectrum) there exists even a constant $\mu < 1$ such that

$$\int_D |\boldsymbol{grad}\, \bar{v}|^2 \, dx < \mu \int_{\Omega^*} |\boldsymbol{grad}\, \bar{v}|^2 \, dx \quad (28)$$

Example 1. Consider in particular that

$$D \subseteq \{ \boldsymbol{x} \in \boldsymbol{R}^3;\ R < r \le R_{\max} \} \quad (29)$$

and $\Omega^* = \{ \boldsymbol{x} \in \boldsymbol{R}^3;\ R < r \}$. In this case $\bar{v}$ may be generally represented by

$$\bar{v} = \sum_{n=0}^{N_{\max}} \left(\frac{R}{r}\right)^{n+1} \bar{v}_n\left(\frac{\boldsymbol{x}}{r}\right), \quad N_{\max} < \infty \quad (30)$$

where $\bar{v}_n$ are the respective Laplace surface spherical harmonics. For μ we then easily obtain

$$\mu \le 1 - (R / R_{\max})^{2N_{\max}+1} \quad (31)$$

As a consequence for functions of band limited spectrum the inequality (25) may be replaced by

$$|\bar{F}(v)| \le M\sqrt{\mu}\, \| v \|_{H_2^{(1)}(\Omega^*)} \quad (32)$$

where, obviously,

$$M \le \sqrt{\mu}\, \| u_{k+1}^{(n,m)} - u_k^{(n,m)} \|_{H_2^{(1)}(\Omega^*)} \tag{33}$$

This immediately yields

$$\| \bar{F} \| \le \mu \| u_{k+1}^{(n,m)} - u_k^{(n,m)} \|_{H_2^{(1)}(\Omega^*)} \tag{34}$$

Return now to the bilinear form $A^*(u,v)$ and assume that $\Omega' = \boldsymbol{R}^3 - \bar{\Omega}^*$ is a star-shaped domain with respect to the origin with Lipschitz' boundary. Under this assumption

$$\| v \|^2_{W_2^{(1)}(\Omega^*)} \le \alpha \int_{\Omega^*} | \boldsymbol{grad}\, v |^2 \, d\boldsymbol{x} \tag{35}$$

for all $v \in W_2^{(1)}(\Omega^*)$, with $\alpha = 5$. The *proof* of this inequality can be found in Holota (1997).

Remark 1. It is not difficult to verify that

$$\| v \| = \left(\int_{\Omega^*} | \boldsymbol{grad}\, v |^2 \, d\boldsymbol{x} \right)^{1/2} \tag{36}$$

has all the characteristic properties of a norm. Thus, considering inequality (35), we can see that $\| v \|_{W_2^{(1)}(\Omega^*)}$ and $\| v \|$ are equivalent norms, i.e.

$$\| v \| \le \| v \|_{W_2^{(1)}(\Omega^*)} \le \sqrt{5}\, \| v \| \tag{37}$$

for all $v \in W_2^{(1)}(\Omega^*)$.

Clearly, it follows from inequality (35) that

$$A^*(u,u) \ge \frac{1}{\alpha} \| u \|^2_{H_2^{(1)}(\Omega^*)} \tag{38}$$

holds for all $u \in H_2^{(1)}(\Omega^*)$, which in the usual terminology means that $A^*(u,v)$ is an elliptic bilinear form. One can also easily verify that

$$A^*(u,v) \le \| u \|_{H_2^{(1)}(\Omega^*)} \| v \|_{H_2^{(1)}(\Omega^*)} \tag{39}$$

The properties of $A^*(u,v)$ and the boundedness of the functional $\bar{F}(v)$ make it now possible to apply Lax-Milgram's theorem which has a key position in the theory of the weak solution, see Necas (1967), Rektorys (1974) or Bers et al. (1964). It allows us to deduce that

$$\begin{aligned} &\| u_{k+2}^{(n,m)} - u_{k+1}^{(n,m)} \|_{H_2^{(1)}(\Omega^*)} \le \alpha \| \bar{F} \| \\ &\le c \| u_{k+1}^{(n,m)} - u_k^{(n,m)} \|_{H_2^{(1)}(\Omega^*)} \end{aligned} \tag{40}$$

where $c = \alpha\mu$. In consequence

$$\begin{aligned} &\| u_{k+2}^{(n,m)} - u_{k+1}^{(n,m)} \|_{H_2^{(1)}(\Omega^*)} \\ &\le c^{m+1} \| u_1^{(n,m)} - u_0^{(n,m)} \|_{H_2^{(1)}(\Omega^*)} \end{aligned} \tag{41}$$

so that for any integer $p > 0$

$$\begin{aligned} &\| u_{k+p}^{(n,m)} - u_{k+1}^{(n,m)} \|_{H_2^{(1)}(\Omega^*)} \\ &\le \frac{c^{m+1}}{1-c} \| u_1^{(n,m)} - u_0^{(n,m)} \|_{H_2^{(1)}(\Omega^*)} \end{aligned} \tag{42}$$

and it follows immediately from this inequality that $\| u_{k+p}^{(n,m)} - u_{k+1}^{(n,m)} \|_{H_2^{(1)}(\Omega^*)} \to 0$ for $k \to \infty$, provided that $c < 1$. This means that $[u_k^{(n,m)}]_{k=0}^{\infty}$ is a Cauchy sequence in $H_2^{(1)}(\Omega^*)$. Recalling finally that $H_2^{(1)}(\Omega^*)$ is a complete space, we are ready to conclude that the sequence $[u_k^{(n,m)}]_{k=0}^{\infty}$ converges to a function $u \in H_2^{(1)}(\Omega^*)$ in the norm $\| . \|_{H_2^{(1)}(\Omega^*)}$.

The problem, however, is that in inequality (35) the parameter α is rather large, which in combination with the condition $c < 1$ is a cause of a considerable limitation for the parameter μ. For this reason we try to deduce the inequality (35) under more restrictive assumptions, but tailored to our particular case. This will be the content of the next section.

4 Ellipticity

For geodetic applications it is important to work with Ω^* being the exterior Ω_{ell} of an ellipsoid of revolution of semi-axes a and b, $a \ge b$. In this case it is natural to use ellipsoidal coordinates u, β, λ. They are related to x_1, x_2, x_3 by the equations

$$\begin{aligned} x_1 &= \sqrt{u^2 + E^2}\, \cos\beta \cos\lambda \\ x_2 &= \sqrt{u^2 + E^2}\, \cos\beta \sin\lambda, \quad x_3 = u \sin\beta \end{aligned} \tag{43}$$

where $E = \sqrt{a^2 - b^2}$, see e.g. Heiskanen and Moritz (1967). Clearly, in the coordinates u, β, λ the boundary $\partial\Omega_{ell}$ of Ω_{ell} is defined by $u = b$. Note also that as usual $e = E/a$ means the first (numerical) eccentricity of the ellipsoid under consideration (i.e. the eccentricity of $\partial\Omega_{ell}$).

In the sequel we will need the functional determinant

$$D = \begin{vmatrix} \frac{\partial x_1}{\partial u} & \frac{\partial x_2}{\partial u} & \frac{\partial x_3}{\partial u} \\ \frac{\partial x_1}{\partial \beta} & \frac{\partial x_2}{\partial \beta} & \frac{\partial x_3}{\partial \beta} \\ \frac{\partial x_1}{\partial \lambda} & \frac{\partial x_2}{\partial \lambda} & \frac{\partial x_3}{\partial \lambda} \end{vmatrix} = -(u^2 + E^2 \sin^2\beta) \cos\beta \tag{44}$$

This result can be verified by means of a direct calculation. Finally, it is also easy to show that

$$r^2 = |x|^2 = u^2 + E^2\cos^2\beta \tag{45}$$

Hence, we can write that

$$\begin{aligned}\int_{\Omega_{ell}} \frac{v^2}{r^2}dx &= \int_{\Omega_{ell}} \frac{v^2}{r^2}|D|\,du\,d\beta\,d\lambda \\ &= \int_{\Omega_{ell}} v^2 \frac{u^2+E^2\sin^2\beta}{u^2+E^2\cos^2\beta}\cos\beta\,du\,d\beta\,d\lambda\end{aligned} \tag{46}$$

and it is obvious that

$$\int_{\Omega_{ell}} \frac{v^2}{r^2}dx \le (1+e'^2)\int_{\Omega_{ell}} v^2\cos\beta\,du\,d\beta\,d\lambda \tag{47}$$

where $e' = E/b$ is the second eccentricity of our ellipsoid of revolution.

Our aim is to compute the integral on the right hand side of Eq. (47). Recalling that $v \in H_2^{(1)}(\Omega_{ell})$ and thus is a harmonic function in Ω_{ell}, we can express it as

$$v(u,\beta,\lambda) = \sum_{n=0}^{\infty}\sum_{m=0}^{n}\frac{Q_{nm}(z)}{Q_{nm}(z_0)}v_{nm}(\beta,\lambda) \tag{48}$$

where Q_{nm} are Legendre's functions of the second kind, v_{nm} are the respective Laplace surface spherical harmonics, while

$$z = i\frac{u}{E}, \quad z_0 = i\frac{b}{E} \quad \text{and} \quad i = \sqrt{-1} \tag{49}$$

is the imaginary unit, cf. e.g. Heiskanen and Moritz (1967) or Hobson (1952). Putting

$$J = \int_{\Omega_{ell}} v^2\cos\beta\,du\,d\beta\,d\lambda \tag{50}$$

and using v as in Eq. (48), we obtain

$$J = \sum_{n=0}^{\infty}\sum_{m=0}^{n}\int_b^{\infty}\left[\frac{Q_{nm}(z)}{Q_{nm}(z_0)}\right]^2 du\int_{\partial S_1} v_{nm}^2\,dS \tag{51}$$

in view of the orthogonality of v_{nm} on the unit sphere.

Hence, our problem is to perform the integration with respect to the variable u. This, however, needs a suitable representation of the function Q_{nm}. It seems that in our case the following formula is the most convenient:

$$\begin{aligned}Q_{nm}(z) &= (-1)^m\frac{2^n n!(n+m)!}{(2n+1)!}(z^2-1)^{-\frac{n+1}{2}} \\ &\times F\left(\frac{n+m+1}{2},\frac{n-m+1}{2},\frac{2n+3}{2};\frac{1}{1-z^2}\right)\end{aligned} \tag{52}$$

Here $F(\alpha,\beta,\gamma;x)$ is the hypergeometric function. The formula coincides with Bateman and Erdély (1965), where for the construction of Q_{nm} Legendre's gamma function $\Gamma(x)$ was used and also with Hobson (1952), where on the contrary Gauss' factorial function $\Pi(x)$ was applied. Thus

$$\begin{aligned}\left[\frac{Q_{nm}(z)}{Q_{nm}(z_0)}\right]^2 &= \left(\frac{z_0^2-1}{z^2-1}\right)^{n+1}k_{nm}(u) \\ &= \left(\frac{b^2+E^2}{u^2+E^2}\right)^{n+1}k_{nm}(u)\end{aligned} \tag{53}$$

where

$$k_{nm}(u) = \left[\frac{F\left(\frac{n+m+1}{2},\frac{n-m+1}{2},\frac{2n+3}{2};\frac{E^2}{u^2+E^2}\right)}{F\left(\frac{n+m+1}{2},\frac{n-m+1}{2},\frac{2n+3}{2};e^2\right)}\right]^2 \tag{54}$$

Passing to the hypergeometric series expression for $F(\alpha,\beta,\gamma;x)$, we have

$$k_{nm}(u) = \left[\frac{1+\frac{(n+m+1)(n-m+1)}{2(2n+3)}\frac{E^2}{u^2+E^2}+\cdots}{1+\frac{(n+m+1)(n-m+1)}{2(2n+3)}e^2+\cdots}\right]^2 \tag{55}$$

and it is immediately clear that

$$\frac{1}{F^2\left(\frac{n+m+1}{2},\frac{n-m+1}{2},\frac{2n+3}{2};e^2\right)} \le k_{nm}(u) \le 1 \tag{56}$$

Hence

$$\left[\frac{Q_{nm}(z)}{Q_{nm}(z_0)}\right]^2 \le \left(\frac{b^2+E^2}{u^2+E^2}\right)^{n+1} \tag{57}$$

Before we return to the integration in Eq. (51) we still prove the following useful lemma.

Lemma. Suppose that $e^2 \le 1/2$ and $u \ge b$ then

$$\frac{b^2+E^2}{u^2+E^2} \le \left(\frac{b+E}{u+E}\right)^2 \tag{58}$$

Proof. Through a direct computation we obtain

$$\frac{b^2+E^2}{u^2+E^2} - \left(\frac{b+E}{u+E}\right)^2 = \frac{2E(u-b)(a^2-b^2-ub)}{(u^2+E^2)(u+E)^2} \tag{59}$$

Inspecting the numerator on the right hand side of Eq. (59), we immediately see that $u-b \ge 0$ and that

$$\begin{aligned}&\max_{u\ge b}[a^2-b^2-ub] = a^2-2b^2 \\ &= a^2(2e^2-1) \le 0\end{aligned} \tag{60}$$

in view of our assumptions, which completes the proof.

Now, using our lemma, we are ready to write

$$J \le \sum_{n=0}^{\infty}\sum_{m=0}^{n}\int_b^{\infty}\left(\frac{b+E}{u+E}\right)^{2n+2} du\int_{\partial S_1} v_{nm}^2\,dS \tag{61}$$

and the integration yields

$$J \le (b+E)\sum_{n=0}^{\infty}\sum_{m=0}^{n}\frac{1}{2n+1}\int_{\partial S_1} v_{nm}^2\, dS \tag{62}$$

Finally, returning to Eq. (47), we have

$$\int_{\Omega_{ell}} \frac{v^2}{r^2}\, dx \le c\sum_{n=0}^{\infty}\sum_{m=0}^{n}\frac{1}{2n+1}\int_{\partial S_1} v_{nm}^2\, dS \tag{63}$$

with

$$c = (1+e'^2)(b+E) = b(1+e'+e'^2+e'^3) \tag{64}$$

In order to interpret the inequality (35) for the domain Ω_{ell} we still have to compute the square of the norm $\|v\|$ given by Eq. (36). Using Green's theorem, we can write

$$\|v\|^2 = \int_{\Omega_{ell}} |\,\boldsymbol{grad}\, v\,|^2\, dx = -\int_{\partial\Omega_{ell}} v\frac{\partial v}{\partial n}\, dS \tag{65}$$

where

$$dS = a\sqrt{b^2 + E^2\sin^2\beta}\,\cos\beta\, d\beta\, d\lambda \tag{66}$$

is the surface element of our ellipsoid of revolution and $\partial/\partial n$ means the derivative in the direction of the outer unit normal $\boldsymbol{n}$ of our ellipsoid of revolution. Recalling the series representation of v as in Eq. (48), we have

$$\frac{\partial v(u,\beta,\lambda)}{\partial n} = \sum_{n=0}^{\infty}\sum_{m=0}^{n} Q_{nm}^{-1}(z_0)\frac{\partial Q_{nm}(z)}{\partial n} v_{nm}(\beta,\lambda) \tag{67}$$

In particular, for $u = b$

$$\frac{\partial Q_{nm}(z_0)}{\partial n} = -\frac{(n+1)\,b}{a\sqrt{b^2 + E^2\sin^2\beta}} Q_{nm}^*(z_0) \tag{68}$$

where

$$Q_{nm}^*(z_0) = Q_{nm}(z_0) + \frac{n-m+1}{z_0\,(n+1)} Q_{n+1,m}(z_0) \tag{69}$$

cf. Holota (2001a,b, 2002a,b). Hence

$$\|v\|^2 = b\sum_{n=0}^{\infty}\sum_{m=0}^{m}(n+1)\frac{Q_{nm}^*(z_0)}{Q_{nm}(z_0)}\int_{\partial S_1} v_{nm}^2\, dS \tag{70}$$

and

$$\frac{Q_{nm}^*(z_0)}{Q_{nm}(z_0)} = 1 - \frac{n-m+1}{n+1}\frac{Q_{n+1,m}(z_0)}{Q_{nm}(z_0)} ie' \tag{71}$$

Moreover, using the representation of Q_{nm} and $Q_{n+1,m}$ by means of hypergeometric functions we arrive at

$$\frac{Q_{n+1,m}(z_0)}{Q_{nm}(z_0)} = -ie\frac{n+m+1}{2n+3} k_{nm}^* \tag{72}$$

where

$$k_{nm}^* = \frac{F\left(\frac{n+m+2}{2}, \frac{n-m+2}{2}, \frac{2n+5}{2}; e^2\right)}{F\left(\frac{n+m+1}{2}, \frac{n-m+1}{2}, \frac{2n+3}{2}; e^2\right)} \tag{73}$$

Passing to the hypergeometric series expression for $F(\alpha,\beta,\gamma;x)$, we have

$$k_{nm}^* = \frac{1+\frac{(n+m+2)(n-m+2)}{2(2n+5)}e^2+\cdots}{1+\frac{(n+m+1)(n-m+1)}{2(2n+3)}e^2+\cdots} \tag{74}$$

Inspecting now the right hand side of this equation, we can show that it is a monotonously decreasing function of n and that $\lim_{n\to\infty} k_{nm}^* = 1$. Therefore,

$$k_{nm}^* \ge \inf_{n\in\mathbb{N}_0}[k_{nm}^*] = 1 \tag{75}$$

Remark 2. For small n we approximately have

$$k_{nm}^* \doteq 1 + \frac{2n^2+8n+2m+7}{2(2n+3)(2n+5)}e^2+\cdots \tag{76}$$

In view of Eq. (75) we can write

$$\frac{Q_{nm}^*(z_0)}{Q_{nm}(z_0)} \ge 1 - \frac{(n+1)^2-m^2}{(n+1)(2n+3)}ee' \ge 1 - \frac{1}{2}ee' \tag{77}$$

so that

$$\|v\|^2 \ge b(1-\frac{1}{2}ee')\sum_{n=0}^{\infty}\sum_{m=0}^{m}(n+1)\int_{\partial S_1} v_{nm}^2\, dS \tag{78}$$

Combing now inequalities (63) and (78), we arrive at

$$\int_{\Omega_{ell}} \frac{v^2}{r^2}\, dx \le \alpha'\,\|v\|^2 \tag{79}$$

where

$$\begin{aligned}\alpha' &= \frac{1+e'+e'^2+e'^3}{1-\frac{1}{2}ee'}\max_{n\in\mathbb{N}_0} f(n)\\ &= \frac{1+e'+e'^2+e'^3}{1-\frac{1}{2}ee'} f(0)\end{aligned} \tag{80}$$

$$f(n) = 1/(2n+1)(n+1) \tag{81}$$

and $\mathbb{N}_0$ is the set of all nonnegative integers. Thus

$$\|v\|^2_{H_2^{(1)}(\Omega_{ell})} \le \alpha \int_{\Omega_{ell}} |\,\boldsymbol{grad}\, v\,|^2\, dx \tag{82}$$

where $\alpha = 1+\alpha' < 2.08$ since $f(0) = 1$,

$$e' = e(1-e^2)^{-1/2} = e + \frac{1}{2}e^3+\cdots \tag{83}$$

as is well-known and $e^2 \doteq 0.006694$ in the Geodetic Reference System 1980, see Moritz (2000). The parameter α is now considerably smaller than in inequality (35).

5 Ellipticity – low degree components absent

Nevertheless, further restrictions lead to still better estimates. To make it clear we will suppose that N is a positive integer and denote by H_N a finite dimensional space of functions which can be represented by

$$v(u,\beta,\lambda)=\sum_{n=0}^{N-1}\sum_{m=0}^{n}\frac{Q_{nm}(z)}{Q_{nm}(z_0)}v_{nm}(\beta,\lambda) \qquad (84)$$

This makes it possible to decompose the space $H_2^{(1)}(\Omega_{ell})$ into two disjoint parts H_N and H^N, i.e. $H_N \cap H^N = 0$, where H^N is a subspace of those functions from $H_2^{(1)}(\Omega_{ell})$ which in general can be represented by

$$v(u,\beta,\lambda)=\sum_{n=N}^{\infty}\sum_{m=0}^{n}\frac{Q_{nm}(z)}{Q_{nm}(z_0)}v_{nm}(\beta,\lambda) \qquad (85)$$

Inspecting the function $f(n)$, we easily deduce that now for any function $v \in H^N$

$$\| v \|^2_{H_2^{(1)}(\Omega_{ell})} \le \alpha \int_{\Omega_{ell}} | \boldsymbol{grad}\, v |^2 \, dx \qquad (86)$$

but with

$$\alpha = 1+\frac{1+e'+e'^2+e'^3}{1-\frac{1}{2}ee'}f(N) \qquad (87)$$

since $f(n)$ is a decreasing function.

Example 2. Supposing that $N=1$, we obtain $\alpha \doteq 1.17$. For $N=5$ we even have $\alpha \doteq 1.02$.

6 Final remark

In a similar way we can also interpret Example 1 for $D \subseteq \{ x \in \boldsymbol{R}^3;\ b < u \le b_{\max} \}$ and $\Omega^* = \Omega_{ell}$. In this case one can show that with some approximation $\mu \le 1-(b/b_{\max})^{2N_{\max}+1}$. Qualitatively, we have a contraction, i.e. $c<1$, for a majority of practical cases, see Sect. 5, but this doesn't mean too much for functions with high frequency components. This will be the object of a new paper in preparation.

Acknowledgements. The work on this paper and its presentation at the 23rd General Assembly of the International Union of Geodesy and Geophysics, Sapporo, June 30 – July 11, 2003 were supported by the Ministry of Education, Youth and Sports of the Czech Republic through Projects No. LA 089 and No. LN 00A005. This support is gratefully acknowledged.

References

Bers L, John F and Schechter M (1964) Partial differential equations. J Wiley and Sons, Inc., New York-London-Sydney

Bateman H and A Erdély A (1965) Higher Transcendental Functions, Volume 1. Russion translation of the English original published by Nauka Publishers, Moscow

Heiskanen WA and Moritz H (1967) Physical geodesy. W.H. Freeman and Company, San Francisco and London

Hobson EW (1952) The theory of spherical and ellipsoidal harmonics. Inostrannaya Literatura Publishers, Moscow. (Russian translation of the English original published by the University Press, Cambridge, 1931)

Holota P (1997) Coerciveness of the linear gravimetric boundary-value problem and a geometrical interpretation. Journal of Geodesy, 71: 640-651

Holota P (2000) Direct method in physical geodesy. In: Schwarz KP (ed.) Geodesy Beyond 2000, The Challenges of the First Decade. IAG Symposia, Vol. 121, Springer, Berlin-Heidelberg-New York etc.: 163-170

Holota P (2001a) Variational methods in the recovery of the gravity field – Galerkin's matrix for an ellipsoidal domain. In: Sideris MG (ed.) Gravity, Geoid and Geodynamics 2000, GGG2000 IAG Intl. Symp., Banff, Alberta, Canada, July 31 - August 4, 2000. Intl. Assoc. of Geodesy Symposia, Vol. 123, Springer, Berlin-Heidelberg-New York etc., 2001: 277-283

Holota P (2001b) Variational methods in the representation of the gravitational potential. Presented at the AROPA workshop, "Institute d'Europe", Castle of Münsbach, Luxembourg, 23-27 October 2001. Cahiers du Centre Européen de Géodynamique et de Séismologie (to appear)

Holota P (2002a) Model refinements in the recovery of the gravity field of the Earth - Variational approach [CD-ROM]. Proc. IAG 2001 Scientific Assembly – Vistas for Geodesy in the New Millennium, 2-7 Sept. 2001, Budapest, Hungary, IAG and Hungarian Acad. of Sciences

Holota P (2002b) Some topics related to the solution of boundary-value problems in geodesy. Presented at the 5th Hotine-Marussi Symposium on Mathematical Geodesy, Matera, Italy, June 17-21, 2002. Intl. Assoc. of Geodesy Symposia, Springer (to appear)

Krarup T (1969) A contribution to the mathematical foundation of physical geodesy. Danish Geod. Inst., Publ. No. 44, Copenhagen

Moritz H (1980) Advanced physical geodesy. H Wichmann Karlsruhe and Abacus Press Tunbridge Wells Kent

Moritz, H. (2000). Geodetic Reference System 1980. The Geodesist's Handbook, Journal of Geodesy, 74: 128-133

Necas J (1967) Les méthodes directes en théorie des équations elliptiques. Academia, Prague

Necas J and Hlavacek I (1981) Mathematical theory of elastic and elasto-plastic bodies: An introduction. Elsevier Sci. Publ. Company, Amterdam-Oxford-New York

Rektorys K (1974) Variacni metody v inzenyrskych problemech a v problemech matematicke fyziky. SNTL Publishers of Technical Literature, Prague 1974; also in English: Variational methods. Reidel Co., Dordrecht-Boston, 1977

Rozanov Y and Sansò F (2003) The analysis of the Neumann and oblique derivative problem. Weak theory. In: Grafarend E, Krumm F and Schwarze VS (eds.) Geodesy - The Challenge of the Third Millennium. Springer, Berlin-Heidelberg-New York etc.: 213-223

Computation of geopotential coefficients from gravity anomalies on the ellipsoid

S. J. Claessens, W. E. Featherstone
Western Australian Centre for Geodesy, Curtin University of Technology, GPO Box U1987, 6845, Perth, WA, Australia
e-mail: claesses@vesta.curtin.edu.au; tel.: +61 8 9266 2218; fax: +61 8 9266 2703

Abstract. One of the most important stages in the computation of a global geopotential model is the computation of the spherical harmonic coefficients from gravity anomalies on the normal ellipsoid. None of the existing methods provides an exact solution, and all show severe shortcomings in the high degrees of the spectrum. In this paper, a new, theoretically exact method is proposed, which is moreover easily applicable up to very high degree and order (2160 and beyond). The solution of the geopotential coefficients is presented as a weighted sum over "spherically approximated" coefficients of equal order, where the gravity anomalies are presumed to reside on a sphere. The weights depend solely upon the degree and order of the coefficient and the definition of the normal ellipsoid and its gravity field. Numerical comparisons with existing methods show substantial differences, especially in the high degrees, which can be explained by the fact that all previous methods are of limited accuracy

Key words. geopotential models – spherical harmonics – ellipsoidal Boundary Value Problem

1 Introduction

The vast majority of geopotential models, describing the Earth's external gravitational potential, (e.g. EGM96 (Lemoine *et al.* 1998) and EIGEN-1S (Reigber *et al.* 2002)) is represented as a series of spherical harmonics, since the Earth's surface, to a first approximation, is a sphere. The Earth is, however, much more accurately described by an ellipsoid of revolution. Therefore, it is desirable and necessary in view of the present-day accuracy requirements to have a method which can compute spherical harmonic coefficients from gravity anomalies that have been downward continued to the ellipsoid.

An exact method to perform a spherical harmonic analysis from gravity anomalies on the ellipsoid is momentarily lacking, but various approximate methods have been developed. The analytical methods can basically be divided into two different approaches: (1) Computation from a summation over "spherically approximated" coefficients, in which the gravity anomalies are assumed to reside on a sphere. This is the standard approach (e.g. Pellinen 1982; Cruz 1986) and was for example utilised in the construction of the geopotential models OSU86. More recently, new variants of this approach have been introduced by Petrovskaya *et al.* (2001) and Sjöberg (2003). (2) Transformation from ellipsoidal harmonic coefficients, computed from gravity anomalies on the ellipsoid, to spherical harmonic coefficients. This procedure was developed by Jekeli (1981, 1988) and extended upon by Rapp and Pavlis (1990) and was used in the construction of the models OSU91 (Rapp *et al.* 1991) and EGM96 (Lemoine *et al.* 1998).

The first approach provides much simpler formulae than the second one. However, all existing methods of the first kind have in common that problems arise at the high degrees (beyond 360). Gleason (1988) shows that Cruz's solution (1986) has severe shortcomings for degrees even below 360, and Sjöberg (2003) acknowledges that his solution is not sufficient for high degrees. Petrovskaya *et al.* (2001) attempt to circumvent this problem by using an elementary transformation of Cruz's basic formula (1986), but the performance of this method beyond degree 360 is thus far unknown. An alternative new approach is suggested here, based on a rigorous extension of Sjöberg's method (2003). This new approach is, contrary to all other approaches thus far, not limited to order e^2 and provides therefore a theoretically exact solution. It is compared against the existing methods of the first kind described above and a comparison with Jekeli's solution (1988) is currently under investigation.

2 Spherical harmonic coefficients

The Earth's disturbing potential (T) can be expanded at the bounding sphere surrounding the Earth's normal ellipsoid as a series of solid spherical harmonics

$$T(r,\theta,\lambda) = \frac{\mu}{a}\sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^{n+1}\sum_{m=-n}^{n}\overline{T}_{nm}\overline{Y}_{nm}(\theta,\lambda) \tag{1}$$

where r, θ and λ are the geocentric distance, colatitude and longitude respectively, μ is the terrestrial

gravitational constant, a is the semi-major axis of the ellipsoid, $\overline{T}_{nm}$ are the dimensionless spherical harmonic coefficients and $\overline{Y}_{nm}$ are the fully normalised spherical harmonics

$$\overline{Y}_{nm}(\theta,\lambda) = \overline{P}_{n|m|}(\cos\theta)\begin{cases}\cos m\lambda & \text{for} \quad m \leq 0\\ \sin m\lambda & \text{for} \quad m > 0\end{cases} \tag{2}$$

where $\overline{P}_{nm}$ are the fully normalised associated Legendre functions. The coefficients $\overline{T}_{nm}$ can be determined from the disturbing potential on the normal ellipsoid (ε) by an equation based on Green's second identity (Sjöberg 1988)

$$\overline{T}_{nm} = \frac{t_n}{4\pi\mu}\int_{\varepsilon}(T\frac{\partial S}{\partial h} - S\frac{\partial T}{\partial h})d\varepsilon \tag{3}$$

where

$$t_n = \frac{1}{(2n+1)a_n} \tag{4}$$

a_n is an arbitrary constant, h is the external normal to the ellipsoid and S is the function

$$S = \left\{\left(\frac{a}{r}\right)^{n+1} + a_n\left(\frac{r}{a}\right)^{n}\right\}\overline{Y}_{nm}(\theta,\lambda) \tag{5}$$

In this representation, it has been assumed that the topographical and atmospherical effects have been removed. The following relationships hold between the ellipsoidal infinitesimal elements and its spherical counterparts (Sjöberg (2003) and Jekeli (1981) respectively)

$$d\varepsilon = \frac{r^2 d\sigma}{\cos\psi} \tag{6}$$

$$\frac{\partial}{\partial h} = \cos\psi\frac{\partial}{\partial r} - \sin\psi\frac{\partial}{r\partial\theta} \tag{7}$$

where σ is the unit sphere, ψ is the difference in angle between the geocentric latitude and geographical latitude and r is the geocentric distance of the computation point at the ellipsoid

$$r = \frac{a\sqrt{1-e^2}}{\sqrt{1-e^2\sin^2\theta}} \tag{8}$$

The "solid" gravity anomaly (Δg) is defined as (e.g. Vaníček *et al.* submitted)

$$\Delta g = -\frac{\partial T}{\partial h} + \frac{1}{\gamma}\frac{\partial\gamma}{\partial h}T \tag{9}$$

where γ is the normal gravity at the computation point. The discrepancies between this formulation and the definition of the "surface" gravity anomaly, which directly follows from the measurements, are of orders higher than e^2. These can be removed from the gravity anomaly in advance, as is for instance shown for the correction term ε_p in Pavlis (1988).

All the existing methods described in the introduction make use of an approximation of Eq. (9) of order e^2. Furthermore, the methods of the first kind rely on Taylor series terms up to order e^2 to make the transformation from the ellipsoid to the sphere (Cruz (1986) also takes the terms of order e^4 and e^6 into account). This is insufficient, as these terms are multiplied by powers of degree n, where the power of n increases for higher terms. This means that for evaluation of high degree coefficients, an extension much higher than the first few terms is needed.

Besides, the approximation depends on the numerical coincidence that the geodynamic parameter m is approximately half the size of e^2 (Moritz 1980, Eq. (39.12)). Hipkin (submitted) argues this approach and in this study, an alternative procedure is followed. Because the normal derivative is present in both the formula for computation of the geopotential coefficients (Eq. (3)) and in the boundary condition (Eq. (9)), the gravity anomaly can be inserted directly into Eq. (3), and with Eqs. (6) and (7) this gives

$$\overline{T}_{nm} = \frac{t_n}{4\pi\mu}\int_{\sigma}\frac{r^2 S}{\cos\psi}\Delta g + \left(r^2\frac{\partial S}{\partial r} - r\tan\psi\frac{\partial S}{\partial\theta} - \frac{r^2 S}{\cos\psi}\frac{1}{\gamma}\frac{\partial\gamma}{\partial h}\right)T d\sigma \tag{10}$$

The integral kernel contains four parts that are either multiplied by the gravity anomaly Δg or the disturbing potential T, namely

$$\frac{r^2 S}{\cos\psi}, \quad r^2\frac{\partial S}{\partial r}, \quad r\tan\psi\frac{\partial S}{\partial\theta} \quad \text{and} \quad \frac{r^2 S}{\cos\psi}\frac{1}{\gamma}\frac{\partial\gamma}{\partial h}$$

These four parts can all be expanded into Taylor series, which will always converge since the flattening of the Earth's ellipsoid is small. The derivation of the Taylor series terms will be performed in section 4. Afterwards, the dependence on the co-latitude θ can be comprised into the spherical harmonics solely, using a new, general relation between associated Legendre functions, derived in the next section.

3 A new general relation among associated Legendre functions of equal degree

From the relation among non-normalised associated Legendre functions of equal degree, given in e.g. Moritz (1980, Eq. 39-76), the following formula can be derived for the fully normalised spherical harmonics

$$\sin^2\theta\overline{Y}_{nm} = \sum_{j=-1}^{1} K_{nm}^{1,j}\overline{Y}_{n-2j,m} \tag{11}$$

where

$$K_{nm}^{1,-1} = \frac{-1}{2n+3}\sqrt{\frac{[(n+1)^2-m^2][(n+2)^2-m^2]}{(2n+1)(2n+5)}} \quad (12)$$

$$K_{nm}^{1,0} = 1 - \frac{2n(n+1)-2m^2-1}{(2n-1)(2n+3)} \quad (13)$$

$$K_{nm}^{1,1} = \frac{-1}{2n-1}\sqrt{\frac{[(n-1)^2-m^2][n^2-m^2]}{(2n-3)(2n+1)}} \quad (14)$$

Through multiple application of Eq. (11), the following new, more general expression can be found

$$\sin^{2i}\theta \overline{Y}_{nm} = \sum_{j=-i}^{i} K_{nm}^{ij} \overline{Y}_{n-2j,m} \quad (15)$$

where the coefficients K_{nm}^{ij} can be computed by

$$K_{nm}^{ij} = \sum_{k=-1}^{1} K_{nm}^{1,k} K_{n-2k,m}^{i-1,j-k} \quad (16)$$

from the initial values given in Eqs. (12 - 14) where it should be noted that all coefficients K_{nm}^{ij} with $n < 0$ and/or $|j| > i$ are zero. Moreover, also the spherical harmonics $\overline{Y}_{nm}$ should be set to zero for $n < 0$.

Furthermore, from another relation among non-normalised associated Legendre functions (Moritz 1980, Eq. 39-46), the following formula can be derived for the fully normalised spherical harmonics

$$\sin\theta\cos\theta\frac{\partial \overline{Y}_{nm}}{\partial\theta} = \sum_{j=-1}^{1} L_{nm}^{j} \overline{Y}_{n-2j,m} \quad (17)$$

where

$$\begin{cases} L_{nm}^{-1} = -nK_{nm}^{1,-1} \\ L_{nm}^{0} = \frac{3}{2}K_{nm}^{1,0} - 1 \\ L_{nm}^{1} = (n+1)K_{nm}^{1,1} \end{cases} \quad (18)$$

It is important to note that both K_{nm}^{ij} and L_{nm}^{j} are symmetric with respect to order m.

4 Taylor series expansions

In order to be able to make use of the new relation in Eq. (15), derived in section 3, the four parts in Eq. (10) need to be expanded in a series of converging Taylor terms. This is possible due to the small eccentricity of the Earth. Firstly, Jekeli (1981) gives

$$\tan\psi = \frac{e^2\sin\theta\cos\theta}{1-e^2\sin^2\theta} \quad (19)$$

Combining Eqs. (5), (8) and (19) gives for the first of the four parts in Eq. (10)

$$\frac{r^2 S}{\cos\psi} = a^2\left\{1-e^2(2-e^2)\sin^2\theta\right\}^{\frac{1}{2}} \times\left\{\left(\frac{b}{a}\right)^{-n+1}(1-e^2\sin^2\theta)^{\frac{n-3}{2}} + a_n\left(\frac{b}{a}\right)^{n+2}(1-e^2\sin^2\theta)^{-\frac{n+4}{2}}\right\}\overline{Y}_{nm} \quad (20)$$

The three parts containing $\sin^2\theta$ can be written as a converging Taylor series (binomial series), due to the small eccentricity of the Earth. Summation and convolution (Cauchy multiplication) of the Taylor series terms eventually yields

$$\frac{r^2 S}{\cos\psi} = \sum_{i=0}^{\infty} A_{in}\sin^{2i}\theta\overline{Y}_{nm} \quad (21)$$

where

$$A_{in} = a^2 e^{2i}\sum_{k=0}^{i}(2-e^2)^{i-k}\binom{\frac{1}{2}}{i-k}\left\{\left(\frac{b}{a}\right)^{-n+1} \times\binom{\frac{n-3}{2}}{k} + a_n\left(\frac{b}{a}\right)^{n+2}\binom{-\frac{n+4}{2}}{k}\right\} \quad (22)$$

A similar procedure can be applied to the other parts of Eq. (10). For the second part, the expression for the derivative of the function S with respect to r was found from Eq. (5) and with Eq. (8) a formula can be found that can be expanded into a Taylor series in the same fashion

$$r^2\frac{\partial S}{\partial r} = \sum_{i=0}^{\infty} B_{in}\sin^{2i}\theta\overline{Y}_{nm} \quad (23)$$

where

$$B_{in} = \frac{b^2}{a}e^{2i}\sum_{k=0}^{i}\binom{\frac{1}{2}}{i-k}\left\{na_n\left(\frac{b}{a}\right)^{n-1} \times\binom{-\frac{n-1}{2}}{k} - (n+1)\left(\frac{b}{a}\right)^{-n-2}\binom{\frac{n+2}{2}}{k}\right\} \quad (24)$$

For the third part, Eq. (8) was first inserted in the expression for the function S (Eq. 5), after which the derivative with respect to co-latitude θ can easily be found. Combining this with Eq. (19), and making use of Eq. (17), yields

$$r\tan\psi\frac{\partial S}{\partial\theta} = \sum_{i=0}^{\infty}(C_{in}\sin^{2i}\theta\overline{Y}_{nm} + D_{in}\sum_{j=-1}^{1} L_{nm}^{j}\sin^{2i}\theta\overline{Y}_{n-2j,m}) \quad (25)$$

where

$$C_{in} = be^{2i+4} \sum_{j=1}^{\min(i,2)} (2i-2j-3) \times \sum_{k=0}^{i} \binom{-\frac{3}{2}}{i-k} \left\{ (n+1)\left(\frac{b}{a}\right)^{-n-1} \times \binom{\frac{n-1}{2}}{k} - na_n \left(\frac{b}{a}\right)^{n} \binom{-\frac{n+2}{2}}{k} \right\} \quad (26)$$

and

$$D_{in} = be^{2i+2} \sum_{k=0}^{i} \binom{-\frac{3}{2}}{i-k} \left\{ \left(\frac{b}{a}\right)^{-n-1} \times \binom{\frac{n+1}{2}}{k} - a_n \left(\frac{b}{a}\right)^{n} \binom{-\frac{n}{2}}{k} \right\} \quad (27)$$

The fourth part involves the normal gravity γ, which can be computed on the ellipsoid by Somigliana's formula (Heiskanen and Moritz 1967, Eq. 2-78)

$$\gamma = \frac{a\gamma_a \cos^2\varphi + b\gamma_b \sin^2\varphi}{\sqrt{a^2 cos^2\varphi + b^2 \sin^2\varphi}} \quad (28)$$

where γ_a and γ_b are the magnitude of normal gravity at the equator and at the poles respectively, and φ is the geographical latitude. Using the well-known relation between geographical latitude and geocentric latitude (e.g. Bomford 1971, Eq. A.47), Eq. (28) can be rewritten into

$$\gamma = \gamma_b (1 - e^2 \sin^2\theta)^{-\frac{1}{2}} \left\{1 - e^2(2-e^2)\sin^2\theta\right\}^{-\frac{1}{2}} \times \left[1 - \left\{1 - \frac{b^3\gamma_a}{a^3\gamma_b}\right\} \sin^2\theta\right] \quad (29)$$

Moreover, from Heiskanen and Moritz (1967, Eqs. 2-79 and 2-80) the following Bruns-type equation can be derived

$$\frac{1}{\gamma}\frac{\partial\gamma}{\partial h} = -\frac{1}{M} - \frac{1}{N} - 2\frac{\omega^2}{\gamma} \quad (30)$$

where ω is the angular velocity of the Earth's rotation and M and N are the principal radii of curvature, given as a function of geocentric co-latitude θ by

$$M = \frac{a^2}{b}(1 - e^2\sin^2\theta)^{-\frac{3}{2}}\left\{1 - e^2(2-e^2)\sin^2\theta\right\}^{\frac{3}{2}} \quad (31)$$

$$N = \frac{a^2}{b}(1 - e^2\sin^2\theta)^{-\frac{1}{2}}\left\{1 - e^2(2-e^2)\sin^2\theta\right\}^{\frac{1}{2}} \quad (32)$$

Combining Eqs. (29-32) the following formula can be found for the fourth and final part

$$\frac{r^2 S}{\cos\psi}\frac{1}{\gamma}\frac{\partial\gamma}{\partial h} = \sum_{i=0}^{\infty} E_{in} \sin^{2i}\theta \overline{Y}_{nm} \quad (33)$$

where

$$E_{in} = -\sum_{k=0}^{i} A_{i-k,n} \left[\frac{b}{a^2} e^{2i} \sum_{j=0}^{k} (2-e^2)^j \times \left\{ \binom{-\frac{3}{2}}{k}\binom{\frac{3}{2}}{k-j} + \binom{-\frac{1}{2}}{k}\binom{\frac{1}{2}}{k-j} \right\} - 2\frac{\omega^2}{\gamma_b} \sum_{j=0}^{k} e^{2(k-j)} \binom{\frac{1}{2}}{k-j} \sum_{l=0}^{k} e^{2(k-l)} (2-e^2)^{k-l} \left(1 - \frac{b^3\gamma_a}{a^3\gamma_b}\right)^{2l} \binom{-1}{l}\binom{\frac{1}{2}}{k-l} \right] \quad (34)$$

5 Computation of geopotential coefficients

Inserting Eqs. (21), (23), (25) and (33) into Eq. (10), followed by insertion of Eq. (15) and interchanging of the order of summation and integration gives

$$\overline{T}_{nm} = \frac{t_n}{4\pi\mu}\left\{ \sum_{i=0}^{\infty} A_{in} \sum_{j=-i}^{i} K_{nm}^{ij} \int_\sigma \Delta g \overline{Y}_{n-2j,m} d\sigma + \sum_{i=0}^{\infty} (B_{in} - C_{in} - E_{in}) \sum_{j=-i}^{i} K_{nm}^{ij} \times \int_\sigma T\overline{Y}_{n-2j,m} d\sigma - \sum_{i=0}^{\infty} D_{in} \sum_{k=-1}^{1} L_{nm}^{k} \times \sum_{j=-i}^{i} K_{n-2k,m}^{ij} \int_\sigma T\overline{Y}_{n-2(j+k),m} d\sigma \right\} \quad (35)$$

The integrals in this expression are equal to those in the spherical approximation, where it holds that

$$\overline{T}^{o}_{nm} = \frac{a}{4\pi\mu}\int_\sigma T\overline{Y}_{nm}(\theta,\lambda)d\sigma = \frac{a^2}{4\pi\mu(n-1)}\int_\sigma \Delta g\overline{Y}_{nm}(\theta,\lambda)d\sigma \quad (36)$$

where use was made of the relationship between the disturbing potential T and the gravity anomaly Δg in the spherical spectral domain (e.g. Rummel and Van Gelderen 1995) After using Eqs. (36) and reshuffling the summation order, the final compact formula

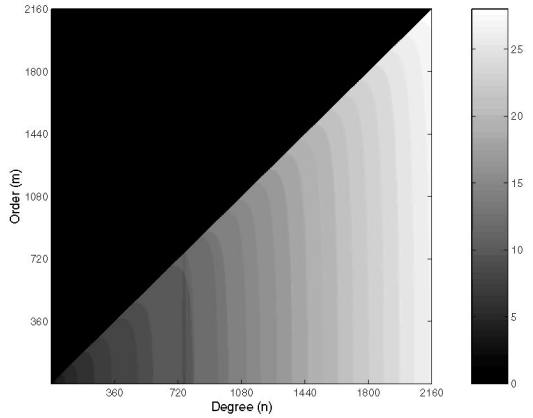


Fig. 1. Convergence rate in the computation of the weights for the case $j = 0$ (summation over i in Eq. (38)): number of terms with an absolute value higher than e^4 for all pairs of degree n and order m

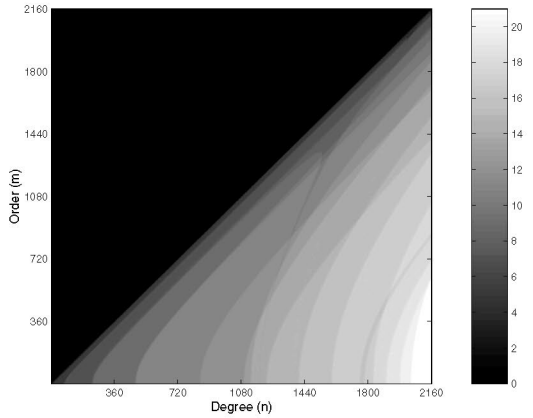


Fig. 2. Convergence rate in the computation of the coefficients (summation over j in Eq. (37)): number of terms with an absolute value higher than e^4 for all pairs of degree n and order m

for the computation of geopotential coefficients from gravity anomalies on the ellipsoid becomes

$$\overline{T}_{nm} = \frac{t_n}{a} \sum_{j=-\infty}^{\infty} M_{nm}^{j} \overline{T}^{o}_{n-2j,m} \qquad (37)$$

where

$$M_{nm}^{j} = \sum_{i=|j|}^{\infty} \left\{ \left(\frac{n-1}{a} A_{in} + B_{in} - C_{in} - E_{in} \right) \times K_{nm}^{ij} - D_{in} \sum_{k=\max(j-i,-1)}^{\min(j+i,1)} L_{nm}^{k} K_{n-2k,m}^{i,j-k} \right\} \qquad (38)$$

Thus, the geopotential coefficients can be obtained from a weighted summation over the spherically approximated coefficients of the same order m. The weights depend solely on the degree and order of the coefficient and the normal ellipsoid and its gravity field. Note that the weighting factors are symmetrical with respect to order m (contrary to the case for a non-normalised spherical expansion), which approximately halves the necessary computation time.

6 Numerical results

Since the first linear eccentricity of the Earth's normal ellipsoid is very small, the series in the final formulae (Eqs. 37 and 38) will converge for every pair of degree n and order m, but for the method to be practically applicable a sufficiently fast convergence rate is necessary. Numerical tests were conducted to derive the convergence rates of both the summation over j in Eq. (37) and the summation over i in Eq. (38). Figures 1 and 2 show the convergence rates for both series, expressed as the number of terms (premultiplied by the factor t_n/a, appearing in Eq. (37)) with an absolute value exceeding e^4. It can be seen that in the summation over i, the number of terms that need to be taken into account for an accurate computation generally with degree n. On the other hand, in the summation over j the trend is that the number of terms increases for an increasing difference between degree n and order m. In both cases, not more than thirty terms need to be taken into account for an accurate computation for even the highest degrees up to 2160.

The numerical results are compared to the existing methods by Cruz (1986), Petrovskaya *et al.* (2001) and Sjöberg (2003), in which the geopotential coefficients are also computed as a weighted sum over "spherically approximated" coefficients. A strategy for a solid comparison with Jekeli's method (1988) is currently under investigation. For simplicity, only the zonal harmonics ($m = 0$) were taken into account and only the weights for $j = -1, 0, 1$ were considered, since Petrovskaya's and Sjöberg's methods only apply these terms. The results of this comparison are shown in Fig. 3. The black lines denote the value of the primary weight ($j = 0$), which defines the relation between spherically approximated coefficients and the final coefficients of the same degree and order. This term is dominant in the low degrees, but its significance decreases in favour of the other terms in the higher degrees. It can be seen that the method presented here is close to Petrovskaya's method (2001) in the lower degrees, but substantial differences occur in the higher degrees. The methods by Cruz (1986) and Sjöberg (2003) are exactly equal to each other, but deviate strongly from the method

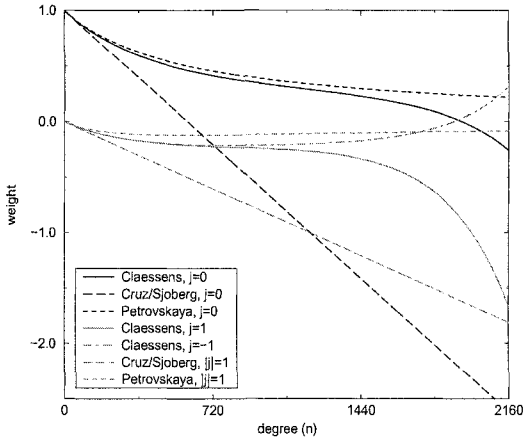


Fig. 3. Comparison between the weight factors of indices $j = -1, 0, 1$ in Cruz's, Sjöberg's and Petrovskaya's method, as well as the method presented here, for zonal harmonics ($m = 0$)

presented here. This can be explained by the insufficient approximation of the boundary condition.

The grey lines in Fig. 3 show the secondary weights ($j = -1$ and $j = 1$). In all existing methods, these two weights are almost equal, but in the new method presented here, the two weights start to diverge significantly from degree ~ 720 onwards. Again, in the low degrees the new method agrees best with Petrovskaya's method, but in the high degrees the differences are very large. These differences can theoretically cause errors in the coefficients in the order of magnitude the original coefficient itself. The influence on the coefficients of the Earth's disturbing potential is currently under investigation.

7 Conclusions

A new, analytical method to compute spherical harmonic geopotential coefficients from gravity anomalies on the ellipsoid has been developed. If evaluated to infinity, this method is theoretically exact under the boundary condition given in Eq. (9), unlike any of the previously employed methods. The method involves an infinite, weighted summation over "spherically approximated" coefficients of equal order, where the weights are computed by an infinite summation as well. Both summations converge sufficiently fast, assuring the method to be easily practically applicable. Comparing the weights used in the new rigorous method to the weights used in existing methods yields substantial differences, especially in the high degrees (beyond degree 360). These differences can be explained by the fact that all existing methods are approximations of insufficient accuracy for computation of high degree coefficients.

References

Bomford G (1971) *Geodesy* (third edition), Oxford University Press, England

Cruz JY (1986) Ellipsoidal Corrections to Potential Coefficients Obtained from Gravity Anomaly Data on the Ellipsoid. Report 371, Department of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio

Gleason DM (1988) Comparing ellipsoidal corrections to the transformation between the geopotential's spherical and ellipsoidal spectrums. Manuscr Geod 13:114-129

Heiskanen WA, Moritz H (1967) Physical Geodesy. W.H. Freeman, San Francisco, California

Hipkin RG (submitted) Ellipsoidal geoid computation. J Geod

Jekeli C (1981) The downward continuation to the Earth's surface of truncated spherical and ellipsoidal harmonic series of the gravity and height anomalies. Report 323, Department of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio

Jekeli C (1988) The exact transformation between ellipsoidal and spherical harmonic expansions. Manuscr Geod 13:106-113

Lemoine FG and 14 other authors (1998) The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM96. NASA/TP-1998-206861, Goddard Space Flight Center, Greenbelt, Maryland

Moritz H (1980) Advanced Physical Geodesy. Herbert Wichmann, Karlsruhe/Abacus Press, Tunbridge Wells

Pavlis NK (1988) Modeling and estimation of a low degree geopotential model from terrestrial gravity data. Report 386, Department of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio

Pellinen LP (1982) Effects of the Earth's ellipticity on solving geodetic boundary value problem. Boll Geod Sci Aff 41:89-103

Petrovskaya MS, Vershkov AN, Pavlis NK (2001) New analytical and numerical approaches for geopotential modeling. J Geod 75: 661-672

Rapp RH, Pavlis NK (1990) The development and analysis of geopotential coefficient models to spherical harmonic degree 360. J Geophys Res 95: 21885 - 21911

Rapp RH, Wang YM, Pavlis NK (1991) The Ohio State 1991 Geopotential and sea surface topography harmonic coefficient models. Report 410, Department of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio

Reigber C and 12 other authors (2002) A high-quality global gravity field model from CHAMP GPS tracking data and accelerometry (EIGEN-1S). Geophysical Research Letters 29:37

Rummel R, Van Gelderen M (1995) Meissl scheme - spectral characteristics of physical geodesy. Manuscr Geod 20:379-385

Sjöberg LE (1988) A new integral approach to geopotential coefficient determinations from terrestrial gravity or satellite altimetry data. Bull Geod 62: 93-101

Sjöberg LE (2003) The ellipsoidal corrections to order e^2 of geopotential coefficients and Stokes' formula. J Geod 77:139-147

Vaníček P, Tenzer R, Sjöberg LE, Martinec Z, Featherstone W (submitted) An analysis of Bouguer gravity anomaly. Geophys J Int

Advanced Harmonic Development of the Earth Tide Generating Potential

S. M. Kudryavtsev
Sternberg Astronomical Institute of Moscow State University, 13 Universitetsky Pr., Moscow, 119992, Russia

Abstract. A harmonic development of the Earth tide generating potential (TGP) by an improved spectral method over two thousands years, 1000-3000, is done. The expansion is directly made to Poisson series where both amplitudes and frequencies of the series' terms are high-degree polynomials of time as opposed to classical Fourier analysis where the terms' amplitudes and frequencies are constants.

The final development of the TGP, named KSM03, includes some 27,000 terms of amplitudes down to the level of $1*10^{-8}$ m^2/s^2. The respective accuracy in calculation of gravity tides at a mid-latitude station is 0.025/0.39 nGal (the r.m.s./maximum error) when comparing with a benchmark gravity tide series numerically computed with use of the most accurate ephemeris DE-405 at every hour within 1600-2200. It exceeds the accuracy of any previously made harmonic development of the TGP in time domain by a factor of at least 3.

Key words: tide generating potential, harmonic development, spectral method

1 Introduction

Doodson (1921) first performed an accurate representation of the Earth tide generating potential (TGP) by harmonic series. Subsequent expansions were done by Cartwright and Tayler (1971), Cartwright and Edden (1973), Büllesfeld (1985), Xi (1987, 1989), and Tamura (1987, 1995). The latest and to date most accurate harmonic developments of the TGP have been made by Hartmann and Wenzel (1994, 1995) and Roosbeek (1996).

In our work we made a new harmonic development of the TGP by employing an improved method of spectral analysis. According to the method the expansion of the TGP is directly made to Poisson series with the terms' amplitudes and arguments being high-degree polynomials of time (as opposed to classical Fourier analysis where the terms' amplitudes are constants and the arguments are linear functions of time). As a source of the planetary and lunar coordinates, the most up-to-date ephemerides DE/LE-405, -406 (Standish, 1998a) have been used. In order to decrease the effect of close frequencies the spectral analysis is done over a two thousand years' interval of time, 1000-3000.

2 Formulation of TGP expansion problem

The classical representation of the Earth TGP generated by external attracting bodies (the Moon, Sun, planets) at an arbitrary point P on the Earth's surface at epoch t is

$$V(t) = \sum_j \mu_j \sum_{n=2}^{\infty} \frac{r^n}{r_j^{n+1}(t)} P_n(\cos\psi_j(t)) \qquad (1)$$

where V is the value of the TGP at P; r is the geocentric distance to P; μ_j, r_j are, respectively, the gravitational parameter and geocentric distance to the j^{th} body; ψ_j is the angle between P and the j^{th} body as seen from the Earth center; P_n is the Legendre polynomial of degree n.

The expression (1) is expanded in our study as

$$V(t) \equiv \sum_{n=2}^{\infty}\sum_{m=0}^{n} V_{nm}(t) = \sum_{n=2}^{\infty}\sum_{m=0}^{n}\left(\frac{r}{R_E}\right)^n \overline{P}_{nm}(\sin\varphi') \times$$

$$[C_{nm}(t)\cos m\theta^{(A)}(t) + S_{nm}(t)\sin m\theta^{(A)}(t)] \qquad (2)$$

where

$$C_{nm}(t) = \frac{1}{2n+1}\sum_j \frac{\mu_j}{R_E}\left(\frac{R_E}{r_j(t)}\right)^{n+1} \times$$

$$\overline{P}_{nm}(\sin\delta_j(t))\cos m\alpha_j^{(A)}(t) \qquad (3)$$

$$S_{nm}(t) = \frac{1}{2n+1}\sum_j \frac{\mu_j}{R_E}\left(\frac{R_E}{r_j(t)}\right)^{n+1}$$

$$\overline{P}_{nm}(\sin\delta_j(t))\sin m\alpha_j^{(A)}(t) \qquad (4)$$

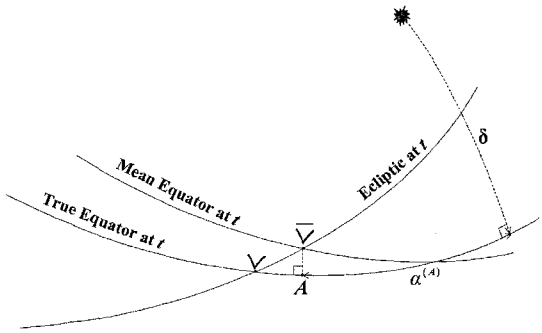


Fig. 1 Spherical coordinates used when developing the TGP

and R_E is the mean Earth equatorial radius; $\alpha_j^{(A)}(t)$, $\delta_j(t)$ are, respectively, the instantaneous right ascension and declination of the j^{th} body referred to the true geoequator of epoch t with an origin point A - that being the projection of the mean equinox of date (see Fig.1); $\theta^{(A)}(t)$ is the local sidereal time at P reckoned from the same point A - so that it is related to the Earth fixed east longitude (from Greenwich) λ of P simply as

$$\theta^{(A)}(t) = \lambda + GMST \tag{5}$$

($GMST$ is Greenwich Mean Sidereal Time defined by a well-known expression by Aoki et al. 1982); φ' is the geocentric latitude of the point P; $\overline{P}_{nm}$ is the normalized associated Legendre function related to the unnormalized one (P_{nm}) as

$$\overline{P}_{nm} = N_{nm} P_{nm} \tag{6}$$

where $N_{nm} = \sqrt{\dfrac{\delta_m (2n+1)(n-m)!}{(n+m)!}}$

and $\delta_m = \begin{cases} 1 \ if \ m = 0 \\ 2 \ if \ m \neq 0. \end{cases}$

The classical expression for the TGP (1) has to be completed by some additional terms reflecting the main effect of Earth's flattening (Wilhelm 1983; Dahlen 1993; Hartmann and Wenzel 1995; Roosbeek 1996) which can be re-written as follows

$$V^{fl}(t) = \frac{r}{R_E}\left\{\overline{P}_{10}(\sin\varphi')C_{10}(t) + \overline{P}_{11}(\sin\varphi')\times \right.$$

$$\left. \left[C_{11}(t)\cos\theta^{(A)}(t) + S_{11}(t)\sin\theta^{(A)}(t)\right]\right\} \tag{7}$$

where

$$C_{10}(t) = \sqrt{\frac{15}{7}}\frac{\bar{J}_2}{R_E}\sum_j \mu_j \left(\frac{R_E}{r_j(t)}\right)^4 \overline{P}_{30}(\sin\delta_j(t)) \tag{8}$$

$$C_{11}(t) = \sqrt{\frac{10}{7}}\frac{\bar{J}_2}{R_E}\sum_j \mu_j \left(\frac{R_E}{r_j(t)}\right)^4 \times$$

$$\overline{P}_{31}(\sin\delta_j(t))\cos\alpha_j^{(A)}(t) \tag{9}$$

$$S_{11}(t) = \sqrt{\frac{10}{7}}\frac{\bar{J}_2}{R_E}\sum_j \mu_j \left(\frac{R_E}{r_j(t)}\right)^4 \times$$

$$\overline{P}_{31}(\sin\delta_j(t))\sin\alpha_j^{(A)}(t). \tag{10}$$

$\bar{J}_2$ is the normalized value for the dynamical form-factor of the Earth ($\bar{J}_2 = J_2 / N_{20}$).

The coefficients $C_{nm}(t)$, $S_{nm}(t)$ contain information about instantaneous positions of the attracting bodies at every epoch t at which one calculates the TGP value $V(t) + V^{fl}(t)$. Angles $\alpha_j^{(A)}(t)$ and $\delta_j(t)$ in expressions (3), (4), (8)-(10) are reckoned along and from the true geoequator of the epoch t, so the relevant values for the coefficients $C_{nm}(t)$, $S_{nm}(t)$ fully take into account the effects of both precision and nutation in obliquity. Because of the choice for the origin point A (see Fig.1) nutation in longitude is not included to the values for $C_{nm}(t)$, $S_{nm}(t)$. Otherwise, one would have to repeatedly take the same effect into account in (5), i.e. substitute there Greenwich True Sidereal Time for Greenwich Mean Sidereal Time, what is more complicated. (When expanding (1) to (2) one gets just differences $\theta^{(A)}(t) - \alpha_j^{(A)}(t)$ between the local sidereal time at P and the right ascension of every perturbing body.)

Having harmonic expansions for $C_{nm}(t)$, $S_{nm}(t)$ one can further calculate the time-dependent values of the TGP at an arbitrary point $P(r, \varphi', \lambda)$ on the Earth's surface by using relations (2), (5) and (7). The tidal acceleration along the Earth radius (or «the gravity tide») is obtained as the radial derivative of the TGP

$$g(t) \equiv \frac{\partial V(t)}{\partial r} = \sum_{n=1}^{\infty} \frac{n}{r} \sum_{m=0}^{n} V_{nm}(t) . \qquad (11)$$

In our work the coefficients $C_{nm}(t)$, $S_{nm}(t)$ have been directly expanded to finite h-order Poisson series of the following form

$$f(t) \approx \sum_{k=1}^{N} \left\{ \left[A^c_{k0} + A^c_{k1} t + ... + A^c_{kr} t^h \right] \cos \omega_k(t) + \right.$$

$$\left. \left[A^s_{k0} + A^s_{k1} t + ... + A^s_{kr} t^h \right] \sin \omega_k(t) \right\} \qquad (12)$$

where $A^c_{k0}, A^c_{k1}, ..., A^s_{kh}$ are constants and $\omega_k(t)$ are some pre-defined arguments which are assumed to be q-degree polynomials of time

$$\omega_k(t) = \nu_k t + \nu_{k2} t^2 + ... + \nu_{kq} t^q . \qquad (13)$$

A detail description of the expansion method is done in Kudryavtsev (2003). The results of applying the method to advanced development of the Earth TGP over two thousand years, 1000-3000, are summarized in the following sections.

3 Harmonic Development of the TGP

First, we calculated numerical values for the coefficients C_{nm}, S_{nm} of the TGP expansion according to (3), (4), (8)-(10) at every six hours within 1000-3000. The latest JPL long-term ephemeris DE/LE-406 was employed as a source of the Moon, Sun and planets coordinates.

When calculating the coefficients we used values for the planetary gravitational parameters from Standish (1998a) and values for $\bar{J}_2$ and R_E from the IERS Conventions (McCarthy and Petit, 2003). (The value for the latter constant which further has to be used in (2) and (7) along with expansions of the coefficients is 6378136.3 m.)

The arguments (13) in expansion (12) were selected as follows. From Simon et al. (1994) we took complete fourth-order polynomial expressions for mean longitude of the ascending node of the Moon Ω, for Delaunay variables D, l', l, F (mean elongation of the Moon from the Sun, mean anomaly of the Sun, mean anomaly of the Moon, and mean longitude of the Moon subtracted by Ω, respectively), and for mean longitudes of Venus, Jupiter, Mars, Saturn and Mercury. (As Hartmann and Wenzel (1994) showed the attraction of other planets is negligible when calculating the TGP.) The set of arguments of the Moon, Sun and planets motion referred to the mean ecliptic and equinox of date was chosen using the latest value of the precession constant. Then we approximately evaluated a spectrum of the tabulated numerical values of C_{nm}, S_{nm} at numerous combinations of multipliers of the arguments' frequencies by using classical FFT procedure. After that, every wave in the spectrum which had a preliminary amplitude exceeding or equal to minimum level 10^{-8} m^2/s^2 was expanded to Poisson series by using the improved technique of spectral analysis developed in (Kudryavtsev, 2003). Analysis proves that only coefficients C_{nm}, S_{nm} of degree $n \le 6$ have amplitudes increasing the chosen minimum level. Maximum degree of the terms' arguments, or parameter q in (13), is equal to 4 in our study (it is the maximum degree of time polynomials for Delaunay variables). The maximum order h of the polynomials for amplitudes in Poisson series (12) is chosen equal to 2. The total number of waves included in the final spectrum of the TGP, named KSM03, is equal to 26,753 - what is the total sum of all N in expansions (12) made for every coefficient.

The complete set of coefficients of KSM03 development of the TGP can be found at http://lnfm1.sai.msu.ru/neb/images/ksm/tgp/coeff.zip. (The description of the data format is done in file http://lnfm1.sai.msu.ru/neb/images/ksm/tgp/readme.doc.)

4 Comparisons and Tests of KSM03

We made a comparison of KSM03 series to the most recent developments of the Earth TGP: HW95 (Hartmann and Wenzel, 1995) and RATGP95 (Roosbeek, 1996). Let us remember that unlike the latter expansions, wave frequencies of KSM03 series are done not in a Earth-fixed reference frame but in a celestial reference frame defined by the true geoequator of date and the projection of the mean equinox of date as the origin of right ascensions (Fig.1). This format for the series explicitly separates two different time arguments to be used in calculating the wave frequencies in a Earth-fixed frame. One of them is T_{eph}, the time argument for JPL planetary/lunar ephemerides (Standish, 1998b), which is close to TDB. It is used when calculating the components of the wave frequencies caused by the Moon, Sun and planets motion, and is exactly the time argument in KSM03 series for coefficients $C_{nm}(t)$, $S_{nm}(t)$ in (3), (4), (8-10). The second time argument is UT1, which is used for calculating another component of the final frequencies – the sidereal time (or Earth rotation) rate (Aoki et al., 1981). In our expansion of the TGP the latter time argument is used in (5) only. UT1 time scale is

much less stable than T_{eph}, and UT1 time can hardly be predicted for long to sufficient accuracy - so it is preferable when coefficients $C_{nm}(t)$, $S_{nm}(t)$ are both calculated and expanded without any use of that time argument. When calculating the TGP and gravity tide at a certain epoch the relevant value for UT1 can be taken from, e.g. IERS publications (or determined from observations) and used in (5) for calculating just terms $m\theta^{(A)}$ in (2) and (7).

However, for the purpose of comparison with the previous TGP developments in frequency domain the major waves from KSM03 expansion were re-calculated to a Earth-fixed reference frame. (The Earth rotation rate was assumed to be uniform.) As our analysis shows the main differences in amplitudes of the major waves in HW95, RATGP95 and KSM03 (see Fig. 2) arise from their either aliasing with or separating from some other small waves of very close frequencies (differing by $\pm 2\dot{p}_s$ where p_s is the mean longitude of solar perigee). However if to merge the waves of such close frequencies then the amplitudes of combined terms become almost identical in the compared three developments. Dehant and Bretagnon (1998) strongly recommend combining the terms of frequencies differing by $\pm 2\dot{p}_s$ in the Earth TGP developments and prove this procedure just leads to a polynomial form of the combined terms' amplitudes. This is what KSM03 development exactly has: the terms of such close frequencies are merged and the terms' amplitudes are second-degree polynomials (let us remember that KSM03 expansion of the TGP is made to Poisson series). As "Report of the WG on theoretical tidal model" (Dehant, 1997) states, the aliasing or separating terms of frequency p_s does not change the result of calculation of the Earth tides although can be important in some other applications, where an analytical expansion might be used.

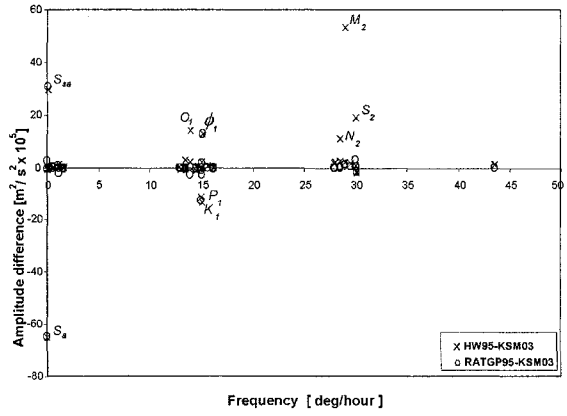


Fig. 2 Differences between amplitudes of major waves in HW95 / RATG95 and KSM03

However, the present analytical developments of the Earth TGP (like RATGP95) do not include many terms of small amplitudes presented in both HW95 and KSM03.

In time domain the accuracy of KSM03 expansion of the Earth TGP has been checked by computation of the gravity tide values (11) at a mid-latitude station. As the latter we choose Black Forest Observatory (BFO) Schiltach: $r = 6366836.9\,m$, $\varphi = 48.3306°\,N$, $\lambda = 8.3300°\,E$ at which Hartmann and Wenzel (1995) and Roosbeek (1996) also computed the tidal gravity by using their expansions of the TGP. First, we calculated the total tidal gravity at that station by means of strict expressions (2)-(4) and (7)-(11) where the Moon, Sun and planets spherical coordinates were computed using the most precise JPL ephemeris DE/LE-405. The gravity tides at BFO were calculated at every hour within the whole time span covered by that ephemeris, 1600-2200. Then we calculated the gravity tides at the same point and at the same set of epochs by using KSM03 expansion of the TGP and compared the results with the exact values. The maximal deviation between the two sets of data at any epoch within the whole time span of six hundred years length does not exceed 0.39 nGal (1 nGal=10^{-11} m/s^2). The corresponding r.m.s. difference between the data over the same interval is less than 0.025 nGal.

Fig.3 shows how the accuracy of calculation of the gravity tide at BFO depends on the number of terms taken from KSM03 series. It is also interesting to estimate how much KSM03 series can be truncated in order to ensure the accuracy of both RATGP95 and HW95 solution (approximately, because KSM03 series have a different format). Thus, the maximum residual 1.23 nGal (the

accuracy of HW95 which includes 12935 terms) is reached when taking 12770 terms from KSM03, and maximum residual 5 nGal (obtained in RATGP95, 6499 terms) is ensured by some 5800 terms from KSM03. Let us remember, that residuals in KSM03 solution are estimated over 1600-2200, while those in HW95 and RATGP95 are done over 1850-2150 and 1987-1993, respectively.

5 Conclusions

• A new harmonic development of the Earth TGP to Poisson series over 1000-3000 (KSM03 solution) is made by means of an improved method of spectral analysis. It includes 26,753 terms of amplitudes down to the level of $1*10^{-8}$ m^2/s^2;

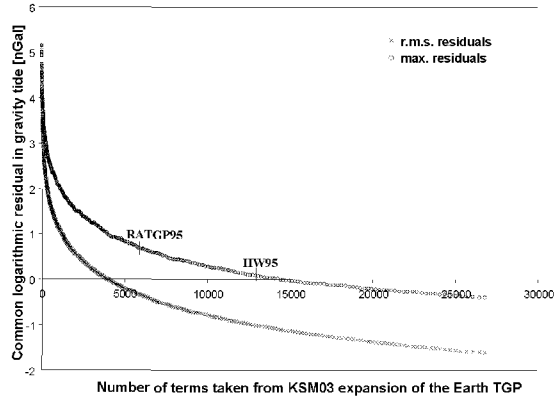


Fig. 3 Residuals between gravity tides calculated by KSM03 expansion of the TGP and those obtained from DE/LE-405 ephemerides (at BFO station over time interval 1600-2200)

• The accuracy of KSM03 in frequency domain is close to that of HW95 and RATGP95 developments. The problem of very close frequencies (differing by $\pm 2\dot{p}_s$) specific to spectral methods (like HW95 and KSM03) is resolved by merging the relevant waves. In this case one gets the amplitudes of the combined waves almost identical in all three developments. Analytical expansions of the TGP (like RATGP95) do not include many waves of small amplitudes presented in both HW95 and KSM03;

• The accuracy of KSM03 in time domain is estimated to be 0.025/0.39 nGal (the r.m.s. / maximum error) when calculating the gravity tides at a mid-latitude station over six hundred years, 1600-2200. It is better than the respective accuracy of any previous TGP development by a factor of at least 3 in time domain and valid over a twice longer interval of time.

Acknowledgements

I am grateful to Drs. V. Dehant and W. Torge for reviewing the manuscript and valuable suggestions and indebted to Charles H. Acton for editing the text. I thank Dr. F. Roosbeek for providing me with his model of the Earth TGP (in HW95 format). Travel grant awarded to me by the IUGG2003 Organizing Committee is sincerely appreciated. The work was supported in part by grant 02-02-16887 from the Russian Foundation for Basic Research.

References

Aoki, S., Guinot, B., Kaplan, G. H., Kinoshita, H., McCarthy, D. D., Seidelmann, P. K. (1982). The new definition of Universal Time. *Astron. Astrophys.*, Vol. 105, pp. 359-361.

Büllesfeld, F. J. (1985) Ein Beitrag zur harmonischen Darstellung des gezeitenerzeugenden Potentials. *Deutsche Geodätische Kommission*, Reihe C, Heft 314, München.

Cartwright, D. E. and Tayler, R. J. (1971) New computation of the tide generating potential. *Geophys. J. R. Astron. Soc.*, Vol. 23, pp. 45-74.

Cartwright, D. E. and Edden, A. C. (1973) Corrected tables of tidal harmonics. *Geophys. J. R. Astron. Soc.*, Vol. 33, pp. 253-264.

Dahlen, F. A. (1993) Effect of the Earth's ellipticity on the lunar tidal potential. *Geophys. J. Int.*, Vol. 113, pp. 250-251.

Dehant, V. (1997) Report of the WG on theoretical tidal model. *Bull. Info. Marées Terrestres*, Vol. 127, pp. 9716-9728.

Dehant, V. and Bretagnon, P. (1998) About the usage of tidal Arguments. *Bull. Info. Marées Terrestres*, Vol. 129, pp. 9946-9952.

Doodson, A. T. (1921) The harmonic development of the tide generating potential. *Proc. R. Soc. Lond. A*, Vol. 100, pp. 305-329.

Hartmann, T. and Wenzel, H.-G. (1994) The harmonic development of the Earth tide generating potential due to the direct effect of the planets. *Geophys. Res. Lett.*, Vol. 21, pp. 1991-1993.

Hartmann, T. and Wenzel, H.-G. (1995) The HW95 tidal potential catalogue. *Geophys. Res. Lett.*, Vol. 22, pp. 3553-3556.

Kudryavtsev, S.M. (2003) Improved Harmonic Development of the Earth Tide Generating Potential. *J. of Geodesy*, in press.

McCarthy, D.D. and Petit G. (eds.) (2003) *IERS Conventions (2000)*, Observatoire de Paris.

Roosbeek, F. (1996) RATGP95: a harmonic development of the tide-generating potential using an analytical method. *Geophys. J. Int.*, Vol. 126, pp. 197-204.

Simon, J. L., Bretagnon, P., Chapront, J., Chapront-Touzé, M., Francou, G., Laskar J. (1994) Numerical expressions for precession formulae and mean elements for the Moon and planets. *Astron. Astrophys.*, Vol. 282, pp. 663-683.

Standish, E.M. and Williams, J.G. (1981) Planetary and lunar ephemerides DE200-LE200. ftp://nav.jpl.nasa.gov/pub/ephem/export/.

Standish, E.M. (1998a) JPL Planetary and Lunar Ephemerides DE405/LE405. *JPL IOM* 312.F-98-048, Pasadena.

Standish, E. M. (1998b) Time scales in the JPL and CfA Ephemerides. *Astron Astrophys.*, Vol. 336, pp. 381-384.

Tamura, Y. (1987) A harmonic development of the tide-generating potential. *Bull. Info. Marées Terrestres*, Vol. 99, pp. 6813-6855.

Tamura, Y. (1995) Additonal terms to the tidal harmonic tables. In: *Proceed. of the 12th Int. Symp. on Earth Tides, Science Press*, Beijing/NewYork, pp. 345-350.

Wilhelm, H. (1983) Earth's flattening on the tidal forcing field. *Geophysics. J. Int.*, Vol. 52, pp. 131-135.

Xi, Q.W. (1987) A new complete development of the tide-generating potential for the epoch J2000.0. *Bull. Info. Marées Terrestres*, Vol. 99, pp. 6766-6812.

Xi, Q.W. (1989) The precision of the development of the tidal generating potential and some explanatory notes. *Bull. Info. Marées Terrestres*, Vol. 105, pp. 7396-7404.

Global Isotropic Isostatic Response Function in Terms of Zonal Spherical Harmonics

Hussein A. Abd-Elmotaal
Civil Engineering Department, Faculty of Engineering,
Minia University, Minia 61111, Egypt
abdelmotaal@lycos.com

Abstract. Assume that the density anomaly is linearly related to the topography by a convolution of the topography and an isotropic kernel function. Hence, it can be shown that the attraction of the compensating masses is also a convolution of the topography and an isotropic isostatic response function, which can be determined by deconvolution. The paper gives the necessary derivation of such a deconvolution by means of global spherical harmonics. A practical determination of the isotropic isostatic response of the earth's crust needs the harmonic analysis of both the topography and the attraction of the compensating masses. To avoid the assumption of an isostatic model, the principle of inverse isostasy has been employed. The harmonic analysis of the Bouguer anomalies is thus a combination of the harmonic analysis of the topographic potential and the already existed global (free-air) reference models. The needed harmonic analysis of the topography has been carried out using different global digital height models. The results show that the isostatic response of the earth's crust derived by inverse isostasy behaves in the same sense as those given by the exact solution of the Vening Meinesz isostatic model.

Keywords. isostasy, inverse problems, response function, density anomaly.

1 Introduction

Dorman and Lewis (1970) have assumed that the compensating density is linearly related to the topography by a convolution of the topographic height and an isotropic kernel function. Thus, the attraction of the compensating masses is a convolution of the topography and an isostatic response function, which can be determined by deconvolution (Moritz, 1990).

A practical determination of the isotropic isostatic response function demands the harmonic analysis of both the topography and the attraction of the compensating masses. The principle of inverse isostasy (forcing the isostatic anomalies to be zero) has been employed. Thus the Bouguer anomalies replace the attraction of the compensating masses. Combining the harmonic analysis of the topographic potential and the free-air reference models perform the harmonic analysis of the Bouguer anomalies.

A practical computation of the isotropic isostatic response function has been carried out using EGM96 model (complete to degree and order 360) and GPM98CR model (complete to degree and order 540) representing the geopotential model. TUG87 ($30' \times 30'$ resolution) and TBASE ($20' \times 20'$ and $30' \times 30'$ resolutions) digital height models have been used for computing the harmonic coefficients of the topography and of the topographic potential. A broad comparison of the isostatic response functions estimated within this investigation has been made. These estimated isostatic response functions are also compared with the Vening Meinesz isostatic model.

It should be noted that the assumption employed here that the density anomaly is linearly related to the topography by a convolution was given originally by Dorman and Lewis (1970). Moritz (1990) has treated this topic in terms of spherical harmonics. Numerical results may be found, e.g., in (Lewis and Dorman, 1970; Bechtel et al., 1987; Hein et al., 1989).

2 Isostatic Response Function

Consider a radial column starting at the earth's center ($r = 0$) till its surface ($r = R$), for which the density anomaly $\Delta\rho$ is a function of the radius vector r, i.e.,

$$\Delta\rho \equiv \Delta\rho(r, \theta, \lambda), \tag{1}$$

where θ and λ denote the spherical co-latitude and longitude, respectively. Let us assume that the density anomaly $\Delta\rho$ is linearly related to the topography h by a convolution (Dorman and Lewis, 1970)

$$\Delta\rho(r', \theta', \lambda') = \iint_{\sigma} h(\theta'', \lambda'') K(r', \psi') \, d\sigma, \tag{2}$$

or symbolically

$$\Delta\rho(r',\theta',\lambda') = h(\theta'',\lambda'') * K(r',\psi'), \qquad (3)$$

where $*$ stands for the spatial convolution and the kernel function $K(r',\psi')$ is isotropic, i.e., it depends *only* on the spherical distance ψ' between the point (r',θ',λ'), for which $\Delta\rho$ is referring, and the point (θ'',λ''), for which the topography h is referring, cf. Fig. 1. The spherical distance ψ' is given by the cosine theorem

$$\cos\psi' = \cos\theta'\cos\theta'' + \sin\theta'\sin\theta''\cos(\lambda''-\lambda'). \qquad (4)$$

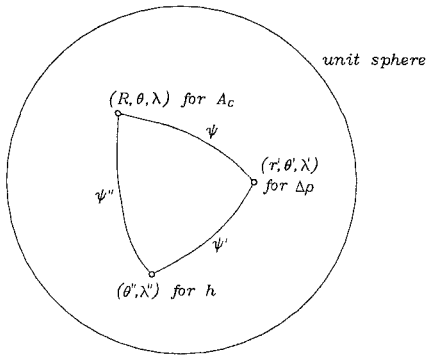


Fig. 1 The used point notations for density anomaly $\Delta\rho$, height h and compensating masses A_C.

The potential of the compensating masses V_C at point (r,θ,λ) is given by Newton's integral as

$$V_C(r,\theta,\lambda) = G\iiint_v \frac{\Delta\rho(r',\theta',\lambda')}{l}\,dv, \qquad (5)$$

where G stands for Newton's gravitational constant, l is the spatial distance between (r,θ,λ) and (r',θ',λ'), given by

$$l^2 = r^2 + r'^2 - 2rr'\cos\psi. \qquad (6)$$

and dv is the spherical volume element, given by

$$dv = r'^2 dr' d\sigma, \qquad (7)$$

and the surface element $d\sigma$ refers to (θ'',λ'').

The attraction of the compensating masses A_C is given by the vertical derivative of the potential of the compensating masses V_C as:

$$A_C(r,\theta,\lambda) = -G\iint_\sigma \int_{r'=0}^{R} \Delta\rho \frac{\partial}{\partial r}\left(\frac{1}{l}\right) r'^2\,dr'd\sigma. \qquad (8)$$

We put

$$D(r',\psi) = -Gr'^2 \frac{\partial}{\partial r}\left(\frac{1}{l}\right). \qquad (9)$$

The dependence of D on r is eliminated by computing A_C at the sea level, for which $r = R$. Thus, the attraction of the compensating masses $\Delta\rho$ is given by

$$A_C(R,\theta,\lambda) = \iint_\sigma \int_{r'=0}^{R} D(r',\psi)\Delta\rho(r',\theta',\lambda')\,dr'd\sigma, \qquad (10)$$

or, symbolically

$$A_C(R,\theta,\lambda) = \int_{r'=0}^{R} D(r',\psi) * \Delta\rho(r',\theta',\lambda')\,dr'. \qquad (11)$$

Substituting (3) into (11) gives

$$A_C(R,\theta,\lambda) = h(\theta'',\lambda'') * \int_{r'=0}^{R} D(r',\psi) * K(r',\psi')\,dr'. \qquad (12)$$

Note that $h(\theta'',\lambda'')$ doesn't depend on r' (see Fig. 1).

The *isostatic response function* F is defined by

$$F(\psi'') = \int_{r'=0}^{R} D(r',\psi) * K(r',\psi')\,dr'. \qquad (13)$$

It depends *only* on the spherical distance between the points (θ,λ) and (θ'',λ'') for reasons of symmetry.

Substituting (13) into (12) gives

$$A_C(R,\theta,\lambda) = h(\theta'',\lambda'') * F(\psi''). \qquad (14)$$

Given the attraction of the compensating masses A_C and the topographic height h, the isostatic response function F can be determined by deconvolution.

3 Deconvolution

To compute the isostatic response function F by deconvolution of (14), we confine ourselves to the global problem, which can be solved by means of spherical harmonics.

Let us expand A_C and F as follows:

$$A_C(R,\theta,\lambda) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}Y_{nm}(\theta,\lambda), \qquad (15)$$

$$F(\psi) = \sum_{n=0}^{\infty} F_n P_n(\cos\psi), \qquad (16)$$

where $P_n(\cos\psi)$ stands for the Legendre polynomial and the spherical harmonics Y_{nm} are given by (Moritz, 1990, p. 195)

$$Y_{nm}(\theta,\lambda) = P_{nm}(\cos\theta)\begin{cases}\cos m\lambda, & m = 0,1,\ldots,n,\\ \sin m\lambda, & m = -1,\ldots,-n,\end{cases} \qquad (17)$$

where $P_{nm}(\cos\theta)$ are the Legendre functions. Since F depends only on ψ, its expansion is purely zonal, cf. (16).

Substituting F from (16) into (14) gives

$$A_C(R,\theta,\lambda) = \sum_{n=0}^{\infty} F_n \iint_{\sigma} h(\theta',\lambda') P_n(\cos\psi)\, d\sigma . \quad (18)$$

By using the well-known integral formula (Heiskanen and Moritz, 1967, eq. 1–71), one can write

$$A_C(R,\theta,\lambda) = 4\pi \sum_{n=0}^{\infty} \frac{F_n}{2n+1} H_n(\theta,\lambda) . \quad (19)$$

We express the Laplace harmonic $H_n(\theta,\lambda)$ of the topography h in terms of the spherical harmonics Y_{nm}

$$H_n(\theta,\lambda) = \sum_{m=-n}^{n} H_{nm} Y_{nm}(\theta,\lambda) . \quad (20)$$

Thus A_C is given by inserting (20) into (19)

$$A_C(R,\theta,\lambda) = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{F_n}{2n+1} H_{nm} Y_{nm}(\theta,\lambda) . \quad (21)$$

Comparing (15) with (21) gives immediately

$$A_{nm} = \frac{4\pi}{2n+1} F_n H_{nm} . \quad (22)$$

Equation (22) is the so-called the *spherical convolution theorem* (Moritz, 1990, p. 251). Equation (22) can be written for F_n as

$$F_n = \frac{2n+1}{4\pi} \frac{A_{nm}}{H_{nm}} , \quad (23)$$

which is *independent* of the order m. This condition should be satisfied by the topography h and the attraction of the compensating masses A_C if the assumption of isotropy is justified. Finally, the isostatic response function is given by (16) with (23).

4 Practical Considerations

One can simply replace the conventional harmonic coefficients appearing in (23) by their fully normalized ones, as they are related by a simple constant (cf. Heiskanen and Moritz, 1967, p. 32, eq. 1–78). Thus, F_n can be given by

$$F_n = \frac{2n+1}{4\pi} \frac{\bar{A}_{nm}}{\bar{H}_{nm}} . \quad (24)$$

Practical computation of F_n by (24) may face a problem of imperfect isotropy. This leads to lose of power of F_n, especially for higher degrees n, if one computes F_n as the mean value over all m, i.e.,

$$F_n = \frac{2n+1}{4\pi} \left(\frac{1}{2n+1} \sum_{m=-n}^{n} \frac{\bar{A}_{nm}}{\bar{H}_{nm}} \right) = \frac{1}{4\pi} \sum_{m=-n}^{n} \frac{\bar{A}_{nm}}{\bar{H}_{nm}} . \quad (25)$$

Alternatively, F_n may be computed by

$$F_n = \frac{2n+1}{4\pi} \frac{\bar{A}_n}{\bar{H}_n} , \quad (26)$$

where $\bar{A}_n$ and $\bar{H}_n$ are given by

$$\bar{A}_n = \sqrt{\sum_{m=-n}^{n} \bar{A}_{nm}^2} , \qquad \bar{H}_n = \sqrt{\sum_{m=-n}^{n} \bar{H}_{nm}^2} . \quad (27)$$

Practical comparison between (25) and (26) will be given in sec. 8.

The summation over n in (16) is infinite. Examining (23) shows that using a certain practical upper limit N_{max}, approximating infinity, will change the values of the isostatic response function (the higher the upper limit, the larger are the values of the isostatic response function). Hence, no direct comparison of isostatic response functions could be made. To solve this problem, we define the *normalized* isostatic response function $\bar{F}(\psi)$ as

$$\bar{F}(\psi) = \frac{F(\psi)}{F(0)} , \quad (28)$$

where $F(0)$ stands for the value of the isostatic response function at the origin ($\psi = 0$).

5 Principle of Inverse Isostasy

A practical determination of the isotropic isostatic response function needs the harmonic analysis of both the topography $\bar{H}_{nm}$ and the attraction of the compensating masses $\bar{A}_{nm}$, cf. (24). Creating compensating masses by means of an isostatic hypothesis already implies an assumption of the earth's isostatic response. Instead, one may apply the principle of inverse isostasy, by which we aim to have zero isostatic anomalies Δg_I (Moritz, 1990, sec. 8.3), i.e.,

$$\Delta g_I = 0 . \quad (29)$$

Equation (29) can alternatively be written in the form

$$A_C = -\Delta g_B , \quad (30)$$

where Δg_B stands for the Bouguer anomalies. Thus, the Bouguer anomalies will be used instead of the attraction of the compensating masses.

6 Harmonic Analysis of the Compensating Masses and Topography

Equation (30) can be written as:

$$A_C = -\Delta g_B = A_T - \Delta g_F . \quad (31)$$

where Δg_F stands for the free-air anomaly and A_T refers to the attraction of the topographic masses. Accordingly, the fully normalized harmonic coefficients of the potential of the compensating masses $\bar{C}_{nm}$ can be defined as

$$\bar{C}_{nm} = \bar{t}_{nm} - \bar{T}_{nm}, \tag{32}$$

where $\bar{t}_{nm}$ are the fully normalized harmonic coefficients of the potential of the topographic masses and $\bar{T}_{nm}$ are the fully normalized harmonic coefficients of the free-air disturbing potential. $\bar{T}_{nm}$ are provided by the global geopotential models (e.g., EGM96). The harmonic analysis of the topographic potential $\bar{t}_{nm}$ will be given in the next section. Finally, the fully normalized harmonic coefficients of the attraction of the compensating masses $\bar{A}_{nm}$ (in gravity anomaly unites) is related to the (unitless) fully normalized harmonic coefficients of the potential of the compensating masses $\bar{C}_{nm}$ by (Heiskanen and Moritz, 1967, p. 89)

$$\bar{A}_{nm} = \frac{GM}{R^2}(n-1)\bar{C}_{nm}, \tag{33}$$

where GM is the geocentric gravitational constant and R is the mean earth's radius.

Expand the topography h in terms of fully normalized harmonic coefficients $\bar{H}_{nm}$

$$h(\theta,\lambda) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \bar{H}_{nm}\bar{Y}_{nm}(\theta,\lambda), \tag{34}$$

where the fully normalized spherical harmonics $\bar{Y}_{nm}$ are given by (analogous to (17))

$$\bar{Y}_{nm}(\theta,\lambda) = \bar{P}_{nm}(\cos\theta)\begin{cases}\cos m\lambda, & m = 0,1,\ldots,n,\\ \sin m\lambda, & m = -1,\ldots,-n,\end{cases} \tag{35}$$

and $\bar{P}_{nm}(\cos\theta)$ are the fully normalized Legendre functions. The fully normalized harmonic coefficients $\bar{H}_{nm}$ are orthogonal, and given by (ibid., p. 31)

$$\bar{H}_{nm} = \frac{1}{4\pi}\iint_{\sigma} h(\theta,\lambda)\bar{Y}_{nm}(\theta,\lambda)\,d\sigma. \tag{36}$$

The practical determination of the fully normalized harmonic coefficients $\bar{H}_{nm}$ is carried out using the HRCOFITR program (Abd-Elmotaal, 2001).

7 Harmonic Analysis of the Topographic Potential

The topographic potential T_T can be written as:

$$T_T(P) = G\iiint_{v}\frac{\rho_Q}{\ell_{PQ}}\,dv_Q, \tag{37}$$

where ρ_Q denotes the density at Q, ℓ_{PQ} is the spatial distance between P & Q and dv_Q is the volume element at Q given by (7). The reciprocal distance can be given by (Heiskanen and Moritz, 1967, p. 33)

$$\frac{1}{\ell_{PQ}} = \sum_{n=0}^{\infty}\frac{r_Q^n}{r_P^{n+1}}P_n(\cos\psi_{PQ}), \tag{38}$$

where r is the modulus of the radius vector, $P_n(\cos\psi)$ is the conventional Legendre polynomial of degree n and ψ_{PQ} is the spherical distance between P and Q.

The Legendre polynomial may be expressed as

$$P_n(\cos\psi_{PQ}) = \frac{1}{2n+1}\sum_{m=-n}^{n}\bar{R}_{nm}(P)\bar{R}_{nm}(Q). \tag{39}$$

Thus the topographic potential can be represented by

$$T_T(P) = G\sum_{n=0}^{\infty}\frac{1}{(2n+1)r_P^{n+1}}\sum_{m=-n}^{n}\bar{R}_{nm}(P)\cdot \left[\iiint_{v}\rho_Q r_Q^n\bar{R}_{nm}(Q)\,dv_Q\right]. \tag{40}$$

The volume integral inside the brackets of (40), confined to the spherical approximation, is given by

$$\iiint_{v} = \frac{R^{n+3}}{n+3}\iint_{\sigma}\rho_Q\left[\left(1+\frac{h_Q}{R}\right)^{n+3}-1\right]\bar{R}_{nm}(Q)\,d\sigma_Q. \tag{41}$$

Hence, the harmonic coefficients of the topographic potential $\bar{t}_{nm}$ and the harmonic series expansion of the topographic potential can be expressed, by inserting (41) into (40), as (Sünkel, 1985)

$$T_T(P) = \frac{GM}{r_P}\sum_{n=0}^{\infty}\left(\frac{R}{r_P}\right)^n\sum_{m=-n}^{n}\bar{t}_{nm}\bar{R}_{nm}(P), \tag{42}$$

$$\bar{t}_{nm} = \frac{R^3}{M(2n+1)(n+3)}\iint_{\sigma}\rho_Q\cdot \left[\left(1+\frac{h_Q}{R}\right)^{n+3}-1\right]\bar{R}_{nm}(Q)\,d\sigma_Q, \tag{43}$$

where M denotes the mass of the earth, given by

$$M \doteq \frac{4\pi R^3}{3}\rho_M, \tag{44}$$

where ρ_M denotes the mean earth's density (Sünkel, 1985)

$$\rho_M = 5.517\ \text{g/cm}^3.$$

For the practical determination of the harmonic coefficients of the topographic potential $\bar{t}_{nm}$, (43)

may be written as

$$\bar{t}_{nm} = \frac{3\Delta\phi\Delta\lambda}{4\pi\rho_M(2n+1)(n+3)} \sum_i^{\phi} \sum_j^{\lambda} \rho_{ij} \left[\left(1+\frac{h_{ij}}{R}\right)^{n+3} - 1 \right] \left\{ \begin{matrix} \cos m\lambda_j \\ \sin m\lambda_j \end{matrix} \right\} \bar{P}_{nm}(\cos\theta_i)\cos\phi_i \,, \qquad (45)$$

where $\sum$ denotes the summation along ϕ and λ, $\Delta\phi$ and $\Delta\lambda$ are the grid sizes of the used digital height model in the latitude and the longitude directions, respectively, and ρ is given by

$$\begin{aligned} \rho &= \rho_o && \text{for } h \geq 0\,, \\ \rho &= \rho_o - \rho_w && \text{for } h < 0\,, \end{aligned} \qquad (46)$$

where ρ_o is the constant density of the topography and ρ_w is the density of the ocean's water.

8 Results

TUG87 ($30' \times 30'$ resolution) and TBASE ($20' \times 20'$ and $30' \times 30'$ resolutions) digital height models have been used in this investigation. EGM96 and GPM98CR models (complete to degree and order 360 and 540, respectively) have been used representing the geopotential model.

Figure 2 shows a comparison of computing F_n by (25) and (26) using TUG87 DHM and EGM96 geopotential model. It shows that for the lower degrees (till around $n = 40$) both curves are nearly coincident. For higher degrees, however, eq. (26) saves the power of F_n with less variations between degree-to-degree values.

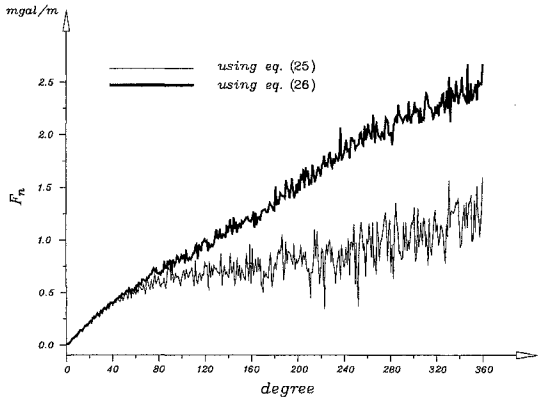


Fig. 2 F_n computed by (25) and (26) using TUG87 DHM and EGM96 geopotential model.

Figure 3 shows F_n computed by (26) using TBASE DHM ($30' \times 30'$ resolution for $N_{max} = 360$ and $20' \times 20'$ resolution for $N_{max} = 540$) and GPM98CR geopotential model. The behaviour of F_n shows, more or less, a linear function, which follows directly the definition of F_n (note the factor (2n+1) appearing in (24) which is an equation of straight line). Figure 3 shows clearly that using a higher value for N_{max}, approximating infinity, will change the values of the isostatic response function F. The normalized isostatic response function $\bar{F}(\psi)$ has been defined to solve this problem.

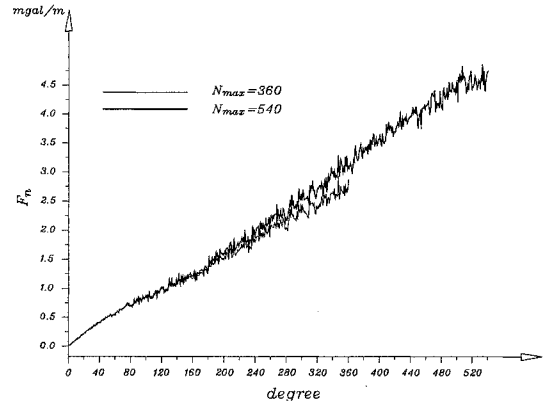


Fig. 3 F_n computed by (26) using TBASE DHM ($30' \times 30'$ resolution for $N_{max} = 360$ and $20' \times 20'$ resolution for $N_{max} = 540$) and GPM98CR geopotential model.

Figure 4 shows the normalized isostatic response functions compared to the Kelvin function kei$x$. Two normalized isostatic response functions are considered here. The first uses EGM96 geopotential model and TUG87 DHM with $N_{max} = 360$. The second uses GPM98CR geopotential model and TBASE DHM ($20' \times 20'$) with $N_{max} = 540$. The Kelvin function kei$x$ (or rather the *scaled* Kelvin function) is the solution of the equation of equilibrium of a thin plate bent by a concentrated load. The Kelvin function kei$x$ then gives the bending curve of the earth's crust subject to the topographic loads. This follows the Vening Meinesz regional isostatic model (Abd-Elmotaal, 1993). For the Kelvin function, a value of the so-called *degree of regionality* l of 20 km has been used. The degree of regionality l is a function of the changeable physical parameters of the earth's crust. Hence, an exact estimation of l is obviously hard, and usually one tries to reasonably assume it.

Figure 4 shows that the normalized isostatic response functions and the Kelvin function kei$x$ give nearly the same values. For the Kelvin function kei$x$, the degree of regionality l may play a role of a fitting parameter. Figure 4 shows that the isostatic response of the earth's crust computed by *direct* isostasy using the most realistic isostatic model (expressed by the

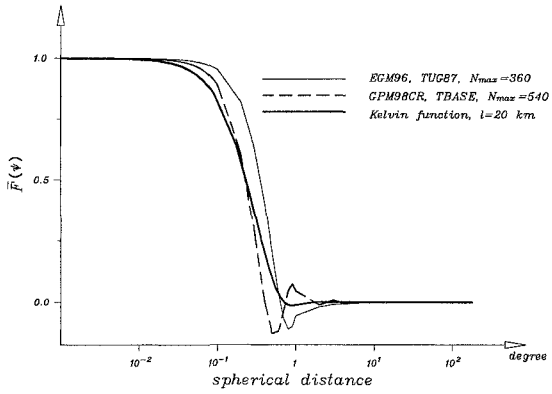


Fig. 4 Normalized isostatic response functions compared to Kelvin function kei$x$.

exact solution of the earth's crust subject to the topographic loads (Abd-Elmotaal, 1993)) matches the isostatic response of the earth's crust derived by *inverse* isostasy.

It should be noted that using GPM98CR geopotential model and TBASE DHM ($30' \times 30'$ resolution) with $N_{max} = 360$ gives a normalized isostatic response function practically coincident with that using EGM96 geopotential model and TUG87 DHM with $N_{max} = 360$. This shows the insignificant effect of the used digital height and geopotential models. The slight disagreement between the two isostatic response functions appearing in Fig. 4 reflects the less accuracy of the used geopotential model with high degree N_{max}.

9 Conclusion

The paper gives the necessary derivation of the isostatic response function by means of spherical harmonics by a deconvolution of the vertical derivative of the isostatic potential. To avoid the assumption of any isostatic model, the principle of inverse isostasy has been employed. The harmonic analysis of the Bouguer anomalies is thus a combination of the harmonic analysis of the topographic potential and the already existed global (free-air) reference models.

A practical computation of the isotropic isostatic response function has been carried out using EGM96 model ($N_{max} = 360$) and GPM98CR model ($N_{max} = 540$) representing the geopotential model. TUG87 ($30' \times 30'$ resolution) and TBASE ($20' \times 20'$ and $30' \times 30'$ resolutions) digital height models have been used for computing the harmonic coefficients of the topography and of the topographic potential. The results show that the normalized isostatic response functions and the Kelvin function kei$x$ give nearly the same values. This shows that the isostatic response of the earth's crust computed from direct isostasy using the most realistic isostatic model (Vening Meinesz isostatic model with Kelvin function kei$x$ as the exact bending curve (Abd-Elmotaal, 1993)) matches the isostatic response of the earth's crust derived by inverse isostasy.

References

Abd-Elmotaal, H. (1993) Vening Meinesz Moho Depths: Traditional, Exact and Approximated, *manuscripta geodaetica*, **18**, 171–181.

Abd-Elmotaal, H. (2001) Enhanced Approach for Harmonic Analysis on the Ellipsoid, *Scientific Bulletin, Ain Shams University, Faculty of Engineering*, **36 (3)**, 53–65.

Bechtel, T.D., Forsyth, D.W. and Swain, C.J. (1987) Mechanisms of Isostatic Compensation in the Vicinity of the East African Rift, Kenya, *Geophysical Journal of Royal Astronomical Society*, **90**, 445–465.

Dorman, L.M. and Lewis, B.T.R. (1970) Experimental Isostasy: 1. Theory of the Determination of the Earth's Isostatic Response to a Concentrated Load, *Journal of Geophysical Research*, **75**, 3357–3365.

Hein, G.W., Eissfeller, B., Ertel, M., Hehl, K., Jacoby, W. and Czerwek, D. (1989) On Gravity Prediction Using Density and Seismic Data, *Preprint, Institute of Astronomical and Physical geodesy, University FAF Munich.*

Heiskanen, W.A. and Moritz, H. (1967) *Physical Geodesy*, Freeman, San Francisco.

Lewis, B.T.R. and Dorman, L.M. (1970) Experimental Isostasy: 2. An Isostatic Model for the USA Derived from Gravity and Topographic Data, *Journal of Geophysical Research*, **75**, 3367–3386.

Moritz, H. (1990) *The Figure of the Earth: Theoretical Geodesy and the Earth's Interior*, Wichmann, Karlsruhe.

Sünkel, H. (1985) An Isostatic Earth Model, *Department of Geodetic Science, The Ohio State University, Columbus, Ohio*, **367**.

Paraná Basin: One isostatic modeling using a Geographic Information System to delineate the main basement structures

M.C.L. Quintas
Departamento de Geomática, Setor de Ciências da Terra
Universidade Federal do Paraná, Centro Politécnico, Curitiba, Paraná, Brasil, 81531-990
e-mail: mquintas@ufpr.br; Tel.: + 55 41 361 3157; Fax: + 55 41 361 3161

Abstract. This paper presents a contribution to the localization of the main basement structures of the Paraná Basin using a Geographic Information System (GIS). This Ordovician-Cretaceous sedimentary basin is intracratonic and located in the S-SE portion of Brazilian territory in South America. The main result of this study, based upon an isostatic modeling approach, was to compute the gravimetric anomaly caused by sediments and igneous rocks that fill the basin, and the "root" created by the isostatic adjustment. Approximately 8,000 gravity stations took part in this study, and through modeling results allowed to identify variations in the basement gravimetric signature, revealing different compositions. Over the years, a great number of gravimetric surveys have been done in the study area and countless data sets were produced. In attempt to summarize and synthesize some of these data sets, a GIS approach was used to create a data bank with topographic, gravimetric and stratigraphic information.

Keywords. Paraná basin – isostatic modeling – GIS

1 Introduction

The stratigraphy of a sedimentary basin is the result of the interaction of a number of different geological processes during time. Studies of the tectonic subsidence determination of a basin by using backstripping techniques have shown that sedimentary factors such as compaction, paleobathymetry and changes in sea level modify the stratigraphy, but their effects are generally small compared with thermal contraction and sedimentary loading. Therefore, the main problem in modeling the stratigraphy of basins is determining the form of the thermal contraction and the nature of the basement's response to sediment loads. In such a context, the concept of isostasy describes the manner in which topographic features on the earth's surface are compensated at depth. Studies indicated that in a number of cases the response of the lithosphere to surface loads can be modeled as that of a thin elastic plate overlying a weak fluid. This model of isostasy, deferred to as the flexure model, is similar to the Airy model, but differs by including the lateral strength of the crust. The effective flexure rigidity is the parameter in the flexure model which characterizes the response of the lithosphere to these loads.

The main focus of this study is the basement structure of the Paraná Basin. An isostatic modeling was applied to approximately 8,000 gravity stations in order to correlate topographic and gravimetric anomalies produced by sub-surface loads. This paper reviews briefly the structure and stratigraphy of the Paraná Basin. This is followed by a discussion on some aspects of the origin of the basin. Finally, it is presented an isostatic study using gravity stations.

2 Structural and Stratigraphic Framework of the Paraná Basin

The Paraná Basin developed on continental crust and filled with sedimentary and volcanic rocks nearly 7,000 m thick that range in age from Ordovician through Cretaceous. Approximately 1,700 m of the basin section consist of Jurassic and Cretaceous volcanic strata, which cover two-thirds of the basin's outcrop area. It comprises an area of 1,400,000 km^2 and extends over parts of Brazil, Paraguay, Uruguay, and Argentina (see Figure 1). The western Argentinean extension of the basin known as Chaco-Paraná, will not be detailed here. The basin has an oval N-S trending shape and the maximum extension is over 1,500 km, with the major axis of the basin lying in N-NE direction, roughly parallel to Paraná River.

The Paraná Basin's basement is composed of continental crust of the Brazilian Platform and several mobile belts of the Brazilian cycle (700-450

Ma). The basin is limited to west by the Asuncion Arch which is possibly an Early Devonian feature (Zalán et al. (1990)). The northwestern margin comprises the Paraguay-Araguaia mobile belt. Towards the north is the western Goiás orogenic belt. Goiania Hinge Zone limits the northeastern margin. Towards the east, the basin is bordered by the ENE-WSW trending Ribeira belt. The southeastern margin is underlying by Dom Feliciano magmatic belt. Exploratory wells and aeromagnetic and gravimetric surveys suggest that the metamorphic basement is affected by two major structural trends: NW-SE and NE-SW. However, the continuity of many of these lineaments remain uncertain, and their possible involvement in the basin development is obscure.

The sediments overlying the crystalline basement of the Paraná Basin accumulated in a suite of three successive basins, which developed in response to varying tectonic settings (Zalán et al. (1990)). The stratigraphy of the basin may further be divided into six major depositional sequences that took place from Ordovician to Cretaceous, each comprising one or more formal lithostratigraphic units. Three of them correspond to Paleozoic transgressive-regressive cycles, and the others are Mesozoic continental sedimentary packages with associated igneous rocks. These depositional sequences are bounded by periods of non-deposition and erosion (Milani (1997)).

Fig. 1. Location of the Paraná Basin.

3 Tectonic History

Controversy has surrounded the understanding of the origin of the Paraná Basin. The basin is crisscrossed by regional normal faults and lineaments, some of which are known to extend into the basement. Previous analyses have suggested the existence of a primarily northeast-southwest (NE-SW) trending Early Paleozoic basin beneath the Mesozoic sedimentary and volcanic sequences that could be an evidence of a central underlying precursor rift. In particular, the Cambro-Ordovician molasses deposits of the Brasiliano cycle may represent remnants of this precursor rift (Cordani et al. (1984)). However, the distribution of these deposits is restricted and the temporal relationship between the beginning of the molasses sedimentation (about 600 Ma) and the first sediments of the Paraná Basin (roughly 440 Ma) does not support a rift basin hypothesis. In addition, the volcanism associated with the Brasiliano cycle is characterized of an intermediate composition of acid and basic components that typifies compression regimes (Zalán et al. (1990)). A unit that is more consistent with a period of crustal stretching is the Late Ordovician/Early Silurian Rio Ivai Formation of the Rio Ivaí super sequence. The isopach map of this unit shows a centrally located, linear depocenter with two subordinate branches, similar in manner to a rift triple junction.

Quintas (1995) calculated subsidence and sedimentation histories for wells on the Paraná Basin and described three major subsidence phases: Silurian/ Devonian (≅ 440 Ma), Permian/ Carboniferous (≅ 300 Ma), and Late Jurassic/ Early Cretaceous (≅ 144 Ma). The subsidence phases are separated by periods of erosion and non-deposition. The first phase of subsidence was responsible for the deposition of Rio Ivai and Rio Paraná super sequences. Pysklywec and Quintas (2000) suggest that the second phase of subsidence may have been caused by the mantle flow mechanism and a combination of contemporaneous processes superimposed on the broader mantle flow-induced epirogeny. The third phase of subsidence coincides with basaltic magmatism associated with the Mesozoic breakup of Gondwana.

4 Isostatic Model

The concept of isostasy describes the process by which the elevation of the earth's surface is adjusted or compensated in response to density changes at depth or to surface loads in order to

bring about uniformity of pressure at some depth, when an area of some size is considered (Dorman and Lewis (1970)).

Basically, there are two types of models of the isostatic compensation of topographic features. In the first model, only the variations in crustal structure are considered: the compensation occurs directly beneath the topography by lateral density variations of crustal (Pratt-Hayford model) or the compensation results from changes in crustal thickness (Airy model). In the second type of model, the flexure model, loads are partially supported by elastic stresses within the lithospheric plate overlying a weak, fluid asthenosphere and compensation occurs on a regional basis (Forsyth (1985)). The flexure model is similar to the Airy model in which surface loads are supported by crustal thickening, but differs by including the lateral strength of the crust. In this model, the response of the plate is characterized by the flexural rigidity.

A common approach to understand the form of isostatic compensation is to study the relationship between topography and the gravity anomalies created by the subsurface mass compensating the topography. Dorman and Lewis (1970) developed a method of quantitatively determination of the state at a geological feature. This method assumes that free air gravity anomalies are caused by topography and its compensation, and attempts to determine a function which, when convolved with topography reproduces the gravity anomaly. The advantage of this method is that this function may be derived from observational data, independent of a particular isostatic model. Once the function has been determined, however, it can be interpreted in terms of simple models of isostasy. The general equation for the gravity effect of a two-layer isostatic model is (Karner and Watts (1982));

$$G_T(K) = 2\pi\gamma\Delta\rho_1 H(K)\exp^{-kd}\left[1-\frac{\Delta\rho_2 R(K)}{\Delta\rho_1 H(K)}\exp^{-kt}\right] \quad (1)$$

$\Delta\rho_1$ is the density contrast across the surface topography, $\Delta\rho_2$ is the density contrast across the root topography, d is the mean elevation, t is the distance between the mean surface and root topography, $H(K)$ is the Fourier transform of the surface topography H(x), $R(K)$ is the Fourier transform of the root topography R(x), γ is the gravitational constant, k is the wave number. The process by which compensation is achieved is determinate by the rheology of the lithosphere. If an elastic rheology is assumed for the lithosphere, then the relation between $R(K)$ and $H(K)$ can be expressed as (Karner and Watts (1982)):

$$R(K) = \frac{\Delta\rho_1}{\Delta\rho_3}\left[1+\frac{k^4 D}{\Delta\rho_3 g}\right]^{-1} H(K) \quad (2)$$

where D is the effective flexural rigidity, $\Delta\rho_3$ is the density contrast between the material underlying the plate and the material infilling the flexural basin. The theoretical admittance $Z(K)$ can be determined for both the Airy and flexure models of isostasy. For example, as $D \rightarrow 0$ and $\Delta\rho 3 \rightarrow \Delta\rho_2$, equation (1) becomes:

$$G_{Airy}(K) = 2\pi\gamma\Delta\rho_1 H(K)e^{-kd}\left[1-e^{-kt}\right] \quad (3)$$

and normalizing by the transform of the surface topography $H(K)$ gives the Airy admittance,

$$Z_{Airy}(K) = 2\pi\gamma\Delta\rho_1 e^{-kd}\left[1-e^{-kt}\right] \quad (4)$$

Similarly, the flexure admittance is obtained by :

$$Z_{flexure}(K) = 2\pi\gamma\Delta\rho_1 e^{-kd}\left\{1-\frac{\Delta\rho_2}{\Delta\rho_3}\left[1+\frac{k^4 D}{\Delta\rho_3 g}\right]^{-1} e^{-kt}\right\} \quad (5)$$

5 Geographic Information System (GIS)

Over the years, many gravimetric studies have been done in this area. In attempt to summarize and synthesize the numerous data sets that have been produced, a GIS approach was used. A database was built using various data sets from different organizations.

By using a GIS, new relationships and correlations were established and the spatial distribution of the attributes were more readily evaluated and quantified. All the data sets were converted into the same projection (UTM, SAD69) and same file format (DBF) and than incorporated into the database. The softwares used were ArcView® v.3.2, and Microsoft Excel®.

6 Modeling Results

This study was carried out in three steps: at first, data collection and database construction using Geographic Information System (GIS); at secondly, statistical analysis to investigate the location and the quality of the data; and finally the isostatic

modeling. Data collected and used to construct the GIS database include: station identification, geographical coordinates, altitude, adjusted gravity value, free air and Bouguer anomalies, terrain correction, photos and station location sketch, date and mean time of the reading, gravimeter type and serial number, sources and quality factors of the adjusted gravity values, and isopach maps of the stratigraphic units. In order to complement the GIS approach, kriging techniques were applied to analyze the spatial patterns.

By applying the admittance function approach to the analysis of free air anomaly and topography profiles, it is possible to investigate the isostatic mechanism operative at the study area. The basin was crisscrossed with 50 profiles (see Figure 1, it shows the position of one profile). The interpretation of the observed admittance function in terms of an isostatic model involves assumptions concerning the rheology and density structure of the basin (Karner and Watts (1982)). It is important to mention that the result of this analysis is an observed admittance function, which reproduces the gravity from the topography and is not tied up to a particular isostatic model. In this study, Airy model was used for the isostatic modeling considering in eq. (4): D = zero, d = the mean elevation changes with the load. The density values are summarized in table 1 as follows.

Table 1. Density values

Type of load	ρ (g/cm^3)
Mantle	3.33
Crust	2.80
Sedimentary-rocks	2.55
Magmatic-rocks	2.87
Rocks	2.67

The isopach maps of the stratigraphic units related to the basin (Quintas, 1995) were used to compute the gravimetric anomalies. The contours map of the computed air gravity anomalies (see Figure 2) shows a low gravimetric between 22°S and 26°S, 53°W and 55°W. This result was expected because of the large load of sedimentary and volcanic rocks in this area, nearly 7,000m thick. Now it is possible to find out the basement influence by estimating the residues between free air anomalies observed and computed. Analyzing the residual contours map (see Figure 3) it is possible to discriminate some trends: a long negative band is detected starting at (48°W, 20°S) and finishing around (54°W, 26°S), expressive

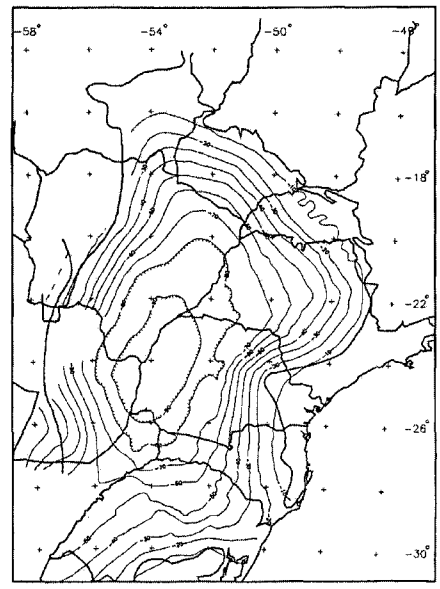


Fig. 2. Gravimetric anomalies produced by subsurface loads.

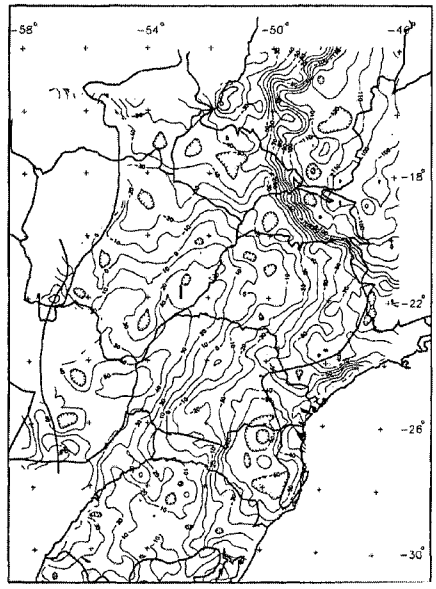


Fig. 3. Gravimetric signature of the main basement structures.
gravimetric gradient is verified at NE of the basin, a trend of positive anomalies starting at (52°W, 20°S) and probably finishing at (58°W, 28°S).

7 Conclusions

The contribution of this study is to establish the gravimetric signature of the main basement structures, as well as the gravimetric anomalies produced by sedimentary and volcanic filling of the Paraná Basin and the "root" created by the isostatic adjustment.

A picture of the basement can be provided by the modeling herein included. The maps shows that the basin is limited to west by the Asuncion Arch. As its northwestern margin, is the Paraguai-Araguaia mobile belt, and according to the modeling may continue beneath the basin sediments and volcanics. Towards the north is the western Goiás orogenic belt. At its northeastern margin, the basin is limited by the Goiânia Hinge Zone. To its east, the basin is bordered by the ENE-WSW trending Ribeira belt, and according to the modeling this mobile belt underlies the basin until 54°W. To its southeastern, the basin is underlying by Dom Feliciano magmatic belt.

Another contribution was the use of GIS that allowed several data sets from different organizations to be organized and managed in a systematic way. Incorporating all these data sets into a common database supports the spatial identification of problematic situations, which will require additional research: areas with not enough gravimetric stations, areas where the altitudes were super or underestimated, and others problems.

Acknowledgement

The author thanks for the release of the data sets: Instituto Astronômico e Geofísico / Universidade de São Paulo (IAG/USP), Instituto Brasileiro de Geografia e Estatística (IBGE), Observatório Nacional (ON), and Universidade Federal do Paraná (UFPR). This work was supported by Conselho Nacional de Ensino e Pesquisa (CNPq – Brazil). The author wishes to thank the anonymous reviewers for critically reviewing this manuscript.

References

Airy, G.B. (1855). On the computation of the effect of the attraction of mountain masses as disturbing the apparent astronomical latitude of stations of geodetic surveys. *Philosophical Transactions of The Royal Society of London*, 145, pp.101-104.

Banks, R.J., R.L. Parker, and S.P. Huestis (1977). Isostatic compensation on a continental scale: local versus regional mechanisms. *Geophysics Journal Research Astr. Society*, 51,pp.431-452.

Barrel, J. (1914). The strength of earth's crust. *Journal of Geology*, 22, n.5, p.441-468.

Cordani, U.G., B.B.B. Neves, R.A. FUCK, R. PORTO, A. Thomas Filho, F.M.B. Cunha (1984). Estudo preliminar de integração do Pré-Cambriano com os eventos tectônicos das bacias sedimentares brasileiras. *Revista Ciência Técnica Petróleo*, 15. Petrobrás, CENPES, CINTEP. 70p.

Cressie, N. (1990). The origins of kriging. *Math. Geol*, 22, n.2, pp.239-252.

Deutsch, C.V., A.G. Journel (1992). Geoestatistical software library and user's guide. Oxford University Press, New York.

Dorman, L.M., B.T.R. Lewis (1970). Experimental Isostasy. 1. Theory of the determination of the earth isostatic response to a concentrated load. *Journal of Geophysical Research*, 75, n.17, pp.3357-3365.

ESRI (Environmental Systems Research Institute) (1998) ArcView 3.2. ESRI, Redlands, California.

Forsyth, D.W. (1985). Subsurface loading and estimates of the flexural rigidity of continental lithosphere. *Journal of Geophysical Research*, 90, n.B14, pp.12.623-12.632.

Heiskanen, W., and F.A. Vening-Meinesz (1958). The earth and its gravity field. McGraw-Hill Book Company, N.Y.

Jeffrey, H. (1959). The earth. 4th ed. Cambridge University Press, London, 420p.

Karner, G.D., and A.B. Watts (1982). On isostasy at Atlantic type continental margins. *Journal of Geophysical Research*, 87, n.B4, pp.2923-2948.

Lewis, B.T.R., and L.M. Dorman (1970). Experimental Isostasy. 2. An isostatic model for the U.S.A. derived from gravity and topographic data. *Journal of Geophysical Research*, 75, n.17, pp.3367-3386.

Milani, E. (1997). Considerações sobre a Bacia do Paraná. Ph.D. Thesis, Universidade Federal do Rio Grande do Sul. 270p.

Pratt, J.H. (1855). On the attraction of the Himalaya mountains and of the elevated regions beyond them. Philosophical *Transactions of The Royal Society of London*,145, pp.53-100.

Pysklywec, R.N., and M.C.L. Quintas (2000). A mantle flow mechanism for the late Paleozoic subsidence of the Paraná Basin. *Journal of Geophysical Research*, 105, n.B7, pp.16359-16370.

Quintas, M.C.L. (1995). O embasamento da Bacia do Paraná: Reconstrução Geofísica de seu arcabouço. Ph.D. Thesis, Universidade de São Paulo, Instituto Astronômico e Geofísico (IAG). 213p.

Quintas, M.C.L., M.S.M. Mantovani, and P.V. Zalan (1999). Contribuição ao estudo da evolução mecânica

da Bacia do Paraná. *Revista Brasileira de Geociências.* 29 (2), pp.217-226.

Walcott, R.I. (1972). Gravity, flexure, and the growth of sedimentary basins at a continental edge. *Geological Society of America Bulletin*, 83,pp.1845-1848.

Watts, A.B., G.D. Karner, and M.S. Steckler (1982). Lithospheric flexure and the evolution of sedimentary basins. *Philosophical Transactions of The Royal Society of London*. Series A,305, pp.249-281.

Zalan, P.V., S. Wolff, J.C. Conceição, A. Marques, M.A.M. Astolfi, I.S. Vieira, V.T. Appi (1990). Bacia do Paraná. In: *Origem e evolução de Bacias Sedimentares.* Petrobrás, Rio de Janeiro, pp.135-164.

Patching local empirical covariance functions – A problem in altimeter data processing

Per Knudsen
Kort & Matrikelstyrelsen, Geodetic Department, Rentemestervej 8, DK-2400 Copenhagen, Denmark.

Abstract. When gravity field modelling is carried out in larger regions a subdivision of the region into smaller areas may be carried out in order to handle the data volume. This may create problems during the estimation since the use of the optimal local empirical covariance function tends to give different results at the borders when the statistical characteristics change from one area to the next. In this paper an approximation to the covariance function that allows spatial variations in both magnitude and in spectral characteristics is presented. The different magnitudes are taken into account by a scaling function. Differences in the spectral characteristics are taken into account by changing the depth to the Bjerhammar sphere, so that it is located at a shallow depth in areas with a rough gravity field. In areas with a smooth gravity field the Bjerhammar sphere is located more deeply.

Introduction

In least squares collocation a covariance function model that agrees with the empirical covariance function should be used in order to obtain optimal estimates. When estimations are carried out in a local area it is advised to use a covariance function that reflects the local statistical characteristics. Such a procedure may create problems in the estimation since the use of the optimal local covariance function may give different results at borders as the statistical characteristics change from one area to the next. In the processing of altimetry for the KMS marine gravity fields such problems exist. In this study, the variability of statistical parameters in the North Atlantic Ocean ($40°<\varphi<80°$, $300°<\lambda<350°$) is analysed. Furthermore, some ways of describing and modelling spatially varying (non-homogeneous) covariance functions are presented.

Modelling covariance functions

The geoid height as well as the free-air gravity anomalies are quantities associated with the anomalous gravity potential T. The geoid height, N, can be expressed in terms of a linear functional (or as in this case linearized functional according to *Bruns' formula*) applied on T (γ is the normal gravity):

$$N = L_N(T) = \frac{T}{\gamma} \tag{1}$$

The free-air gravity anomalies are associated with T through

$$\Delta g = L_{\Delta g}(T) = -\frac{\partial T}{\partial r} - 2\frac{T}{r} \tag{2}$$

Now T is expanded into fully normalized spherical harmonic functions on the surface of a sphere with a radius R (Heiskanen & Moritz, 1967). That is

$$T = \frac{GM}{R}\sum_{i=2}^{\infty}\left(\frac{R}{r}\right)^{i+1} \sum_{j=0}^{i}\left[\overline{C}_{ij}\cos j\lambda + \overline{D}_{ij}\sin j\lambda\right]\overline{\mathrm{P}}_{ij}(\sin\phi) \tag{3}$$

where $T=T(r,\varphi,\lambda)$ and r,φ,λ are spherical coordinates (radius, latitude, and longitude respectively). C_{ij} and D_{ij} are real coefficients and P_{ij} are associated Legendre's functions of degree i and order j (bars denote that the harmonics are fully normalized). Using this spherical harmonic expansion, eq.(3), an expression for geoid heights, eq.(1), is obtained by multiplying the coefficients by $1/\gamma$. If the coefficients are multiplied by $(i\text{-}1)/R$ then an expression for gravity anomalies is obtained.

The covariance values are obtained using the kernel functions. The kernel associated with the gravity field is derived using eq.(3) and some *a-priori* variances. The covariance between T in the points $P(\varphi,\lambda)$ and $Q(\varphi',\lambda')$ is expressed as

$$K(P,Q) = \sum_{i=2}^{\infty}\left(\frac{R^2}{r_P r_Q}\right)^{i+1}\sigma_i^{TT}\,\mathrm{P}_i(\cos\psi) \tag{4}$$

where $\sigma_i^{TT} = (2i+1)\overline{\sigma}_i^{TT}$ are degree variances and ψ is the spherical distance between P and Q. Hence, eq.(4) only depends on the distance between P and Q and neither on their locations nor on their azimuth (i.e. a homogeneous and isotropic kernel).

Expressions associated with geoid heights and gravity anomalies are obtained by applying the respective functionals on $K(P,Q)$, e.g. $C_{NN}=L_N(L_N(K(P,Q)))$ (more on collocation by Moritz, 1980, Sansò, 1986, Tscherning, 1986). Then

$$C_{NN} = \sum_{i=2}^{\infty}\left(\frac{1}{\gamma}\right)^2\left(\frac{R^2}{r_P r_Q}\right)^{i+1}\sigma_i^{TT}\,\mathrm{P}_i(\cos\psi) \tag{5}$$

$$C_{\Delta g\Delta g} = \sum_{i=2}^{\infty}\left(\frac{i-1}{R}\right)^2\left(\frac{R^2}{r_P r_Q}\right)^{i+1}\sigma_i^{TT}\,\mathrm{P}_i(\cos\psi) \tag{6}$$

$$C_{N\Delta g} = \sum_{i=2}^{\infty}\left(\frac{i-1}{\gamma R}\right)\left(\frac{R^2}{r_P r_Q}\right)^{i+1}\sigma_i^{TT}\,\mathrm{P}_i(\cos\psi) \tag{7}$$

The modelling of the covariance function associated with the gravity field is described in Knudsen (1987, 1988). This technique has been applied using empirical covariance values calculated from marine gravity data reduced using a set of spherical harmonic coefficients to degree 360. As degree variance model a Tscherning/Rapp model (Tscherning & Rapp, 1974) was used. This expression has the advantage that the kernel can be evaluated using a closed expression instead of the infinite sum. The model is

$$\sigma_i^{TT} = \begin{cases} \varepsilon_i & i = 2,\ldots,360 \\ \dfrac{A}{(i-1)(i-2)(i+4)}\left(\dfrac{R_B^2}{R^2}\right)^{i+1} & i = 361,\ldots. \end{cases} \tag{8}$$

where A is a constant in $[m^4/s^4]$, R_B is the radius of the Bjerhammar sphere in [m], and ε_i are the error degree variances associated with the spherical harmonic expansion of the reference model.

Such a local empirical covariance function may have been derived using data in local areas of a size of a few degrees. By experience it has been found that an area of 2 by 2 degrees is suitable for calculation of the empirical covariances when a reference model to harmonic degree and order 360 has been subtracted from the observations. Smaller areas usually give noisy and inconsistent results.

A computation of empirical covariance functions in small local areas covering a larger region such as the North Atlantic results in a number of covariance functions having very different characteristics. First of all the magnitude of the gravity field varies substantially from place to place. Furthermore, the spectral contents of the gravity field as expressed through the 50% correlation length, vary.

As mentioned in the introduction the use of different covariance functions in neighbouring areas give different estimates along the borders, even if quite large overlap regions have been used. To overcome such problems caused by the discontinuities in the statistical parameters an attempt is made in this study to derive some (approximate) expressions for covariance functions that vary continuously from one area to another.

Analysis and Case Studies

In this first step an attempt is made to derive an expression for covariance functions that vary continuously from one area to another where only the variances are considered. In a second attempt, variations in the spectral characteristics as expressed by changes in the height are considered. In both cases it is assumed that a spherical harmonic expansion complete to degree and order 360 has been subtracted. The associated error degree variances are ignored.

Case I: Scaling

In this case variations in the variances of the residual gravity field at different locations are considered. The variances are derived over local areas of extents that are large enough to contain wavelengths corresponding to harmonic degree 360 and higher as wavelengths shorter than about 1 degree can be contained fully in areas of 2 by 2 degrees. Furthermore, it is assumed that the local spectral characteristics in the different areas are similar.

Then a smooth surface, $s(\varphi,\lambda)$, is constructed so that it represents the actual standard deviations of the residual gravity field at specific locations. The standard deviations may have been computed in 2 by 2 degree areas. Since the local spectral characteristics have been assumed to be similar, and only hereby, the gravity field may be approximated by a homogenised function T' multiplied by the surface $s(\varphi,\lambda)$. Subsequently, the "scaled gravity field" may be expanded into spherical harmonic functions. That is

$$\begin{aligned} T &\approx L_s\left(T'(\phi,\lambda,r)\right) \\ &= s(\phi,\lambda)T'(\phi,\lambda,r) \\ &= s(\phi,\lambda)\frac{GM}{R}\sum_{i=2}^{\infty}\left(\frac{R}{r}\right)^{i+1} \\ &\quad \sum_{j=0}^{i}\left[\overline{C'}_{ij}\cos j\lambda + \overline{D'}_{ij}\sin j\lambda\right]\overline{\mathrm{P}}_{ij}(\sin\phi) \end{aligned} \tag{9}$$

where L_s. is a functional multiplying the function T' expressed in terms of a spherical harmonic expansion where C'_{ij} and D'_{ij} are fully normalized harmonic coefficients, so that the result is a good approximation of T. The approximation may be good when the spectral content of the surface $s(\varphi,\lambda)$ is characterized by wavelengths sufficiently long to avoid that the local spectral characteristics of T and T' become different. As an example $s(\varphi,\lambda)$ may contain wavelengths longer than 2

degree which corresponds to harmonic degree 90, when T contain shorter wavelengths associated with harmonic degree 360 and higher.

Now T' has been homogenized, so that its local statistical characteristics are similar everywhere in the region or even on the sphere. Hence, a regional or global empirical covariance function is a good approximation to the local empirical covariance functions. Subsequently, the covariance function will be determined for T' and related to T by applying the scaling functional introduced in eq. (9) in points P and Q. That is

$$K(P,Q) \approx L_s^P\left(L_s^Q\left(\sum_{i=2}^{\infty}\left(\frac{R^2}{r_P r_Q}\right)^{i+1}\sigma_i^{T'T'}\mathrm{P}_i(\cos\psi)\right)\right) \qquad (10)$$

where the degree variance associated with the homogenized field can be modelled by eq. (8). Covariances associated with other gravity field related quantities can be obtained by applying their respective functionals additionally.

An example from the North Atlantic Ocean is shown in Figure 1, 2, and 3. Figure 1 shows the KMS marine free-air gravity field in the North Atlantic Ocean ($40°<\varphi<80°$, $300°<\lambda<350°$) relative to the EGM96 geopotential model (Lemoine et al., 1998). The effects of the Mid-Atlantic ridge clearly show up in the residual gravity field. It is evident that the residual gravity field here and in some other parts of the region is quite rough. In other parts of the region the residual gravity is smooth. These local differences in the magnitude of the residual gravity field are displayed in Figure 2 showing the standard deviations of the residual gravity field in cells of 2° by 2°. Again, the pattern of the Mid-Atlantic ridge shows up clearly. Actually, the standard deviation shown in Figure 2 represent the surface $s(\varphi,\lambda)$ in eq.(9) and when the residual gravity field is scaled by this function then T' in eq.(9) is obtained. This homogenized residual gravity field is shown in Figure 3.

Case II: Upward continuation

In this case variations in the spectral characteristics as expressed by changes in the height, or more correctly in the radius, is considered. The motivation for this approach is that the spectral characteristics of the gravity field changes with height and, as an example, it becomes smoother when it is continuated upward. Then a level function or surface at varying height $h(\varphi,\lambda)$, is constructed, so that T mapped onto this surface will have similar local spectral characteristics in different local areas. That is

$$\begin{aligned}\tau(R,\phi,\lambda) &\approx T(R+h,\phi,\lambda)\\ &= \frac{GM}{R}\sum_{i=2}^{\infty}\left(\frac{R}{R+h}\right)^{i+1}\\ &\sum_{j=0}^{i}\left[\overline{C}_{ij}\cos j\lambda+\overline{D}_{ij}\sin j\lambda\right]\overline{\mathrm{P}}_{ij}(\sin\phi)\\ &= \frac{GM}{R}\sum_{i=2}^{\infty}\\ &\sum_{j=0}^{i}\left[\overline{C}'_{ij}\cos j\lambda+\overline{D}'_{ij}\sin j\lambda\right]\overline{\mathrm{P}}_{ij}(\sin\phi)\end{aligned} \qquad (11)$$

where C'_{ij} and D'_{ij} now are fully normalized harmonic coefficients of the spherical harmonic expansion of τ.

Using τ an empirical covariance function can be estimated that will represent the local spectral characteristics everywhere in the region on the surface $R+h$. Then the degree variances can be modelled by eq.(8). To obtain a covariance function associated with T and related quantities the covariances associated with τ should be continuated back from $R+h$ to R. That is

$$\sigma_i^{TT} = \left(\frac{(R+h_P)(R+h_Q)}{R^2}\right)^{i+1}\sigma_i^{\tau\tau} \qquad (12)$$

Combining the results of these two cases a modified covariance function may be derived that changes from place to place to describe the local statistical properties of the gravity field. Still based on the original kernel eq. (4) to utilize the functional relations to gravity anomalies and geoid heights, eq. (5-7), a non-homogeneous covariance function may be obtained using a modified degree variance model such as

$$\sigma_i^{TT} \approx \begin{cases}\varepsilon_i & i = 2,\ldots,360\\ \dfrac{s_P s_Q A}{(i-1)(i-2)(i+4)}\left(\dfrac{(R_B+h_P)(R_B+h_Q)}{R^2}\right)^{i+1} & i = 361,\ldots.\end{cases} \qquad (13)$$

where the location dependencies expressed in eq. (10) and (12) have been introduced in the original degree variance model in eq. (8). In eq. (12) R has been replaced by R_B with negligible effects prior to this. Then the surface h is modifying the radius of the Bjerhammar sphere.

Following the example in the North Atlantic the depths to the Bjerhammar sphere have been estimated in cells of 2° by 2° using the local empirical covariance values. This was done cell by cell, by relating the correlation length (defined as the lag where the covariance is 50% of the variance) to the depth of the Bjerhammar sphere using eq.(6) and eq.(8). The result

shows that the depth to the Bjerhammer sphere is smallest closest to the Mid-Atlantic ridge and that it is located deeper going away from the ridge. Hence, the Bjerghammar sphere appear to be shallow in areas characterised with a rough gravity field and deeply located in areas having a smooth field.

Discussion

With this study some of the problems in using local empirical covariance functions have been addressed. Edge effects caused by discontinuities in the statistical parameters at the borders of neighbouring areas may be avoided if non-homogeneous covariance functions are used. Eq.(13) is an approximation to the covariance function that allows spatial variations in both magnitude and in spectral characteristics. The different magnitudes are taken into account by a scaling function. Differences in the spectral characteristics are taken into account by changing the depth to the Bjerhammer sphere, so that it is located at a shallow depth in areas with a rough gravity field. In areas with a smooth gravity field the Bjerhammer sphere is located more deeply.

The expression (eq.13) is an approximation that may not be very good unless some specific criteria are fulfilled. As mentioned it is believed to be important that the approximation is good only when the spectral contents of the surfaces $s(\varphi,\lambda)$ and $h(\varphi,\lambda)$ are characterized by long wavelengths compared to the spectral characteristics of the residual gravity field where a high degree harmonic expansion has been removed. As an example, $s(\varphi,\lambda)$ may contain wavelengths longer than 2 degrees which correspond to harmonic degree 90, when T contain shorter wavelengths associated with harmonic degree 360 and higher.

Further analyses are needed to verify that the expression may be used as a covariance function. An answer to the question whether it is a covariance function at all should be found. In the meantime the principles will be tested to see if the global altimetric gravity fields computed by Andersen & Knudsen (e.g., 1998) will be improved.

References

Andersen, O. B. and P. Knudsen. Global Marine Gravity Field from the ERS-1 and GEOSAT Geodetic Mission Altimetry. J. Geophys. Res, 103(C4), pp. 8129-8137, 1998.

Heiskanen, W. A., and Moritz, H.: Physical Geodesy, W. H. Freeman, San Francisco, 1967.

Knudsen, P.: Estimation and Modelling of the Local Empirical Covariance Function using gravity and satellite altimeter data. Bulletin Géodésique, Vol. 61, 145-160, 1987.

Knudsen, P.: Determination of local empirical covariance functions from residual terrain reduced altimeter data. Reports of the Dep. of Geodetic Science and Surveying no. 395, The Ohio State University, Columbus, 1988.

Lemoine, F.G., S.C. Kenyon, J.K. Factor, R.G. Trimmer, N.K. Pavlis, D.S. Chinn, C.M. Cox, S.M. Klosko, S.B. Luthcke, M.H. Torrence, Y.M. Wang, R.G. Williamson, r.H. Rapp, and T.R. Olson, The development of the joint NASA/GSFC and the National Imagery and Mapping Agency (NIMA) geopotential models, EGM96, NASA, TP-1998-206861, 575 pp, July 1998.

Moritz, H.: Advanced Physical Geodesy. Herbert Wichmann Verlag, Karlsruhe, 1980.

Sansò, F.: Statistical Methods in Physical Geodesy. In: Sünkel, H.: Mathematical and Numerical Techniques in Physical Geodesy. Lecture Notes in Earth Sciences, Vol. 7, 49-155, Springer-Verlag, 1986.

Tscherning, C.C.: Functional Methods for Gravity Field Approximation. In: Sünkel, H.: Mathematical and Numerical Techniques in Physical Geodesy. Lecture Notes in Earth Sciences, Vol. 7, 3-47, Springer-Verlag, 1986.

Tscherning, C.C., and R.H. Rapp: Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations, and Deflections of the Vertical Implied by Anomaly Degree Variances. Report no. 208, Dept. of Geodetic Science and Surveying, The Ohio State University, Columbus, 1974.

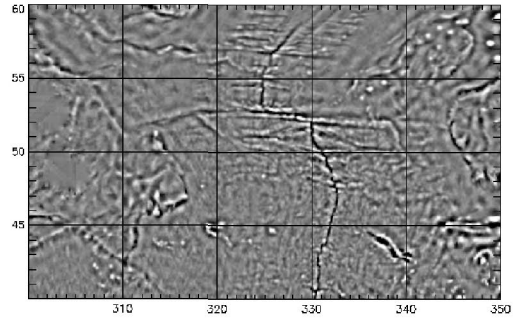


Figure 1. Free-air gravity field in the North Atlantic relative to EGM96. Scale from –40 to 40 mgal.

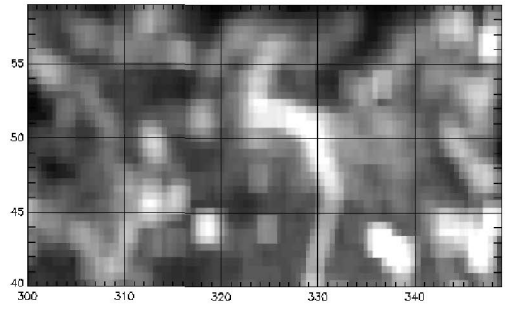


Figure 2. Standard deviation of the free-air gravity field in the North Atlantic relative to EGM96 computed in cells of 2° by 2°. Scale from 0 to 20 mgal.

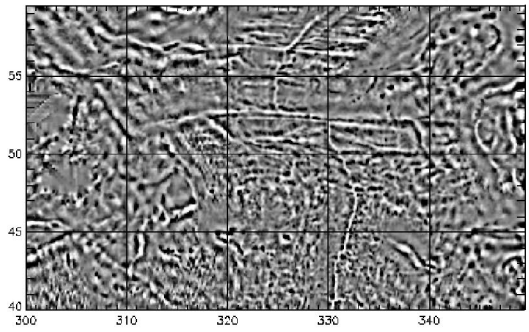


Figure 3. Scales/homogenized free-air gravity field in the North Atlantic relative to EGM96. Scale from –2 to 2..

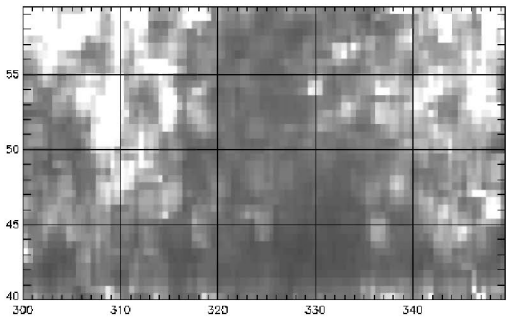


Figure 4. Depths to Bjerhammar sphere estimated from empirical covariance function computed in cells of 2° by 2° covering the North Atlantic. Scale from 0 km to 20 km.

Asymptotic Dislocation Theory of a Spherical Earth for Calculating Co-seismic Displacement

Wenke SUN and Shuhei OKUBO
Earthquake Research Institute, University of Tokyo, Tokyo, Japan
sunw@eri.u-tokyo.ac.jp

Abstract. This study presents an asymptotic theory for calculating displacements excited by a point dislocation in a spherically symmetric earth model as an approximation of the dislocation theory for a spherical earth model (Sun et al., 1996), employing the reciprocity theorem (Okubo, 1993) and asymptotic solutions of an elastic earth to tidal, load and shear forces (Okubo, 1988). This theory is mathematically simple and physically reasonable since it reflects the earth's sphericity and radial structure. Comparison of the asymptotic results with both the exact results and the corresponding flat-earth results shows that for any distances the exact results are approximated better by the asymptotic results than by the flat-earth results. Especially, the asymptotic theory is obviously superior to the flat-earth theory due to a big sphericity effect for the tensile opening on a vertical fault. Owing to its mathematical simplicity, this theory can be applied easily to calculate co-seismic displacements. The validity of the theory for a stratified earth needs to be further tested and discussed.

Key words. Dislocation theory, asymptotic solution, co-seismic displacement, earthquake, spherical earth

1. Introduction

To study the Earth's responses to seismic dislocation, the earth is often simplified into one of three simple geometric models: a half-space, a homogeneous sphere, or a stratified sphere. For the two spherical models, self-gravitation is usually postulated. For these earth models, corresponding dislocation theories have been presented.

The half-space earth model is the simplest; it assumes the earth to be a half–space filled with a homogeneous elastic medium. It is easy to derive mathematical expressions for seismic responses of the medium for such a simple earth model. Many scientists (e.g., Steketee, 1958; Maruyama, 1964; Okada, 1985) presented analytical expressions for calculating surface displacement, tilt, and strain due to various dislocations buried in the half-space medium. Okubo (1991, 1992) proposed expressions in closed form to describe potential and gravity changes due to dislocations. Due to their mathematical simplicity, these dislocation theories have been widely applied up to the present day to study or inverse seismic faults. However, the validity of these theories is strictly limited to a near field because Earth's curvature and radial heterogeneity are ignored.

A homogeneous spherical model with no self-gravitation is obviously superior to half-space since it includes earth curvature. Efforts to develop formations for such an earth model were advanced through numerous studies (among them are Ben-Menahem and Singh, 1968; Smylie and Mansinha, 1971). These studies revealed that earth's curvature effects are negligible for shallow events, while vertical layering may have considerable effects on deformation fields. However, Sun and Okubo's (2002) recent study comparing the discrepancy between a half-space and a homogeneous sphere and between a homogeneous sphere and a stratified sphere indicates that both curvature and vertical layering have significant effects on co-seismic deformation. Saito (1967) presented a theory to calculate amplitudes of free oscillations caused by a point source in a spherically symmetric sphere. He expressed his results in terms of normal mode solutions and source functions.

Stratified spherical models, such as the 1066A model (Gilbert and Dziewonski, 1975), are the most realistic. Sun and Okubo (1993) and Sun et al. (1996) presented theories to calculate co-seismic displacements and gravity changes in spherically symmetric earth models. Okubo (1993) proposed a reciprocity theorem for connecting solutions of dislocation and tidal, shear, and load deformations. He found that deformation on the earth's surface caused by dislocations at source radius $r = r_s$ can be expressed simply by a linear combination of tide, shear, and load deformations at $r = r_s$. This theoretical result provides a useful tool for the current study. Piersanti et al. (1995, 1997) and

Sabadini et al. (1995) studied displacement and rates induced by a dislocation in viscoelastic, stratified earth models, accounting for sphericity and self-gravitation using a self-consistent approach.

The above studies concerning different earth models indicate that different theories have different advantages and disadvantages. For the half-space earth model, corresponding theories are mathematically simple and can be used easily in an application. However, the disadvantage of the theories is that the earth model is physically too simple to reflect sphericity and vertical structure of the earth. On the other hand, theories for spherical earth models are physically better, but are mathematically tedious due not only to numerical integration and summation, but also complicated numerical tables (Sun and Okubo 1993). Therefore, to overcome the disadvantages of the two cases, the current study seeks a new way to express the excited displacement field by considering its approximation—asymptotic solutions. Merits of this new theory are that it is mathematically simple and physically reasonable.

Applying the reciprocity theorem (Okubo, 1993) and asymptotic solutions of the elastic earth to tide, shear, and load sources (Okubo, 1988) plus the powerful tool of infinite sums of Legendre functions, the excited displacement field by dislocations buried in a spherically symmetric earth model can be easily derived in an analytical form. This research is limited to an elastic earth model, but it can be extended easily to a viscoelastic one if necessary. The objective of this study is computation of co-seismic displacements, but it can be extended easily to other geodetic deformations, such as gravity, tilt, or strain, in the same way.

2. Basic Equations of Elastic Responses of a Sphere to a Point Dislocation

We start with a dislocation model as shown in Fig. 1, which is defined at radial distance r_s on an infinitesimal fault dS by slip vector **v**, normal **n** (**v** and **n** are unit vectors), slip angle λ, and dip angle δ in the coordinate system ($\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3$); unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are taken in the equatorial plane in the direction of longitude $\varphi = 0$ and $\pi/2$, respectively, and $\mathbf{e}_3$ along the polar axis **r**. The relative movement of the two fault sides is defined as ($U/2$)-(-$U/2$) =U. Therefore, displacement vector $U\mathbf{v}$ and normal **n** of the source can be expressed as

$$\begin{aligned} U\mathbf{v} &= U(v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3) \\ \mathbf{n} &= n_1\mathbf{e}_1 + n_2\mathbf{e}_2 + n_3\mathbf{e}_3 \end{aligned} \tag{1}$$

Note that for a tensile opening, the slip vector and the normal become the same, i.e., **v**=**n**.

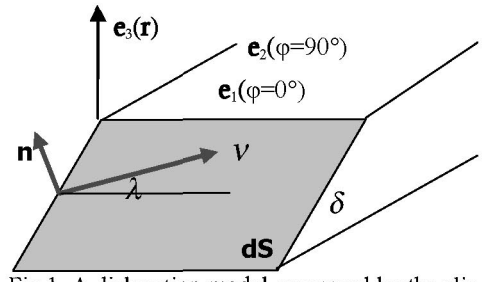


Fig.1. A dislocation model expressed by the slip vector **v**, normal **n**, slip angle λ, and dip angle δ in the coordinate system ($\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3$).

If dislocation occurs in a spherical earth, such as a homogenous sphere or a SNREI (spherically symmetric, non-rotating, perfectly elastic and isotropic) earth, the excited displacement **u** at (r,θ,φ) can be described as

$$\begin{aligned} \mathbf{u}(r,\theta,\varphi) = \sum_{n,m,i,j} \Big[& y_{1,m}^{n,ij}(r)\mathbf{e}_r + y_{3,m}^{n,ij}(r) \\ & \cdot\left(\mathbf{e}_\theta \frac{\partial}{\partial\theta} + \mathbf{e}_\varphi \frac{1}{\sin\theta}\frac{\partial}{\partial\varphi}\right) + y_{1,m}^{t,n,ij}(r) \\ & \cdot\left(\mathbf{e}_\theta \frac{1}{\sin\theta}\frac{\partial}{\partial\varphi} - \mathbf{e}_\varphi \frac{\partial}{\partial\theta}\right)\Big] Y_n^m \cdot v_i n_j \frac{UdS}{a^2} \end{aligned} \tag{2}$$

where $(\mathbf{e}_r,\mathbf{e}_\theta,\mathbf{e}_\varphi)$ are unit vectors in spherical coordinates, $Y_n^m(\cos\theta)$ is the surface spherical harmonic with degree n and order m, a is the earth radius. Note that the factor a^2 in (2) comes from a normalization of y-symbols, in which variables $y_{k,m}^{n,ij}(a)$ and $y_{k,m}^{t,n,ij}(a)$ in (2) become dimensionless.

Once the solutions $y_{k,m}^{n,ij}(a)$ and $y_{k,m}^{t,n,ij}(a)$ are obtained numerically, the excited displacement field in (2) can be calculated. Since i=1, 2, 3 and j=1, 2, 3, the combination of i and j indicates that the total number of solutions of all y's should be nine. However, due to symmetry of source functions (Saito, 1967), and the intrinsic symmetry within fault geometry, the number of independent solutions of $y_{k,m}^{n,ij}(a)$ and $y_{k,m}^{t,n,ij}(a)$ is four. This means that if any four independent solutions are obtained, other solutions among the nine are also easily obtained. In this study, we choose ($y_{k,m}^{n,12}$, $y_{k,m}^{n,32}$, $y_{k,m}^{n,22}$, $y_{k,m}^{n,33}$) (and

the corresponding toroidal solution) for four independent solutions. They are excited by a vertical strike-slip, a vertical dip-slip, a horizontal opening along a vertical fault, and a vertical opening along a horizontal fault, respectively.

3. Expressions of the Four Independent Asymptotic Displacements

In order to consider each component of displacement **u**, Eq. (2) can be rewritten as

$$\mathbf{u}(a,\theta,\varphi)=\sum\left[u_r^{ij}\mathbf{e}_r+\left(u_\theta^{s,ij}+u_\theta^{t,ij}\right)\mathbf{e}_\theta+\left(u_\varphi^{s,ij}+u_\varphi^{t,ij}\right)\mathbf{e}_\varphi\right]\cdot\nu_i n_j\frac{UdS}{a^2} \tag{3}$$

where the components, u_r^{ij}, $u_\theta^{s,ij}$, and so on, are given by Eq. (2), as functions of $y_{k,m}^{n,ij}(a)$ and $y_{k,m}^{t,n,ij}(a)$. Since the solutions of $y_{k,m}^{n,ij}(a)$ and $y_{k,m}^{t,n,ij}(a)$ can be expressed in the form of asymptotic series (Okubo, 1988), summations in (3) can be easily obtained in an analytical form. Omitting the tedious mathematical work, and let $c=\cos\theta$, $s=\sin\theta$, $w=\sqrt{1-2\varepsilon\cos\theta+\varepsilon^2}$ and $\varepsilon=r_s/a$, the final expressions of these displacement components can be obtained as

$$\begin{aligned}u_r^{12}(a,\theta,\varphi)&=2\sin 2\varphi\cdot\left\{y_{230}^{12}\left[\frac{2c}{w^3}-\frac{1}{w^5}\left(2c-\varepsilon(3+c^2-2\varepsilon c)\right)\right]\right.\\&+y_{231}^{12}\frac{1}{\varepsilon}\left[\frac{2}{s^2}-1-\frac{1-\varepsilon c}{w^3}+\frac{2c(\varepsilon-c)}{s^2w}\right]+y_{232}^{12}\left[\frac{\varepsilon s^2}{2}\right.\\&\cdot\left(\frac{1}{w^3}+\frac{2\left(1+w^{-1}\right)+w^{-3}\varepsilon c}{1+w-\varepsilon c}+\frac{\varepsilon c\left(1+w^{-1}\right)^2}{(1+w-\varepsilon c)^2}\right)-\frac{c}{\varepsilon^2s^2}\\&\left.\left.-\frac{1}{2\varepsilon^2s^2w}\cdot\left(\varepsilon\left(1+c^2\right)-2c+w^{-2}\varepsilon s^2(\varepsilon c-1)\right)\right]\right\}\end{aligned} \tag{4a}$$

$$\begin{aligned}u_\theta^{s,12}(a,\theta,\varphi)&=2\sin 2\varphi\left\{y_{330}^{12}\left[\frac{2s}{w^3}-\frac{3}{w^5}\varepsilon s\right.\right.\\&\cdot(\varepsilon-c)-\frac{2}{\varepsilon s^3}(2-s^2)\left(c+\frac{\varepsilon-c}{w}\right)\\&\left.+\frac{2c}{\varepsilon s}\left(-1+\frac{1}{w}-\frac{\varepsilon(\varepsilon-c)}{w^3}\right)\right]+y_{331}^{12}\frac{1}{\varepsilon}\\&\left.\left[\frac{\varepsilon s}{w^3}+\frac{2(2-s^2)}{\varepsilon s^3}(1-\varepsilon c-w)-\frac{2c}{s}\left(1-\frac{1}{w}\right)\right]\right\}\end{aligned} \tag{4b}$$

$$\begin{aligned}u_\varphi^{s,12}(a,\theta,\varphi)&=\frac{4\cos 2\varphi}{s}\left\{y_{330}^{12}\frac{1}{\varepsilon}\right.\\&\cdot\left[1-\frac{1}{w}+\frac{\varepsilon(\varepsilon-c)}{w^3}+\frac{2c}{s^2}\left(c+\frac{\varepsilon-c}{w}\right)\right]\\&+y_{331}^{12}\frac{1}{\varepsilon s}\left(1-\frac{1}{w}-\frac{2c}{\varepsilon s^2}(1-\varepsilon c-w)\right)+\left(y_{331}^{12}+y_{332}^{12}\right)\\&\left[\left(\frac{c}{\varepsilon^2s^3}+\frac{1}{2\varepsilon^2s^3w}\left(\varepsilon\left(1+c^2\right)-2c+\frac{\varepsilon s^2}{w^2}(\varepsilon c-1)\right)\right)\right.\\&+\frac{\varepsilon s}{2}\left(\frac{1}{w^3}+\frac{2\left(1+w^{-1}\right)+w^{-3}\varepsilon c}{1+w-\varepsilon c}+\frac{\varepsilon c\left(1+w^{-1}\right)^2}{(1+w-\varepsilon c)^2}\right)\\&\left.\left.-\frac{1}{\varepsilon s}\left(1-\frac{1}{w}+\frac{\varepsilon(\varepsilon-c)}{w^3}+\frac{2c}{s^2}\left(c+\frac{\varepsilon-c}{w}\right)\right)\right]\right\}\end{aligned} \tag{4c}$$

$$\begin{aligned}u_\theta^{t,12}(a,\theta,\varphi)&=-4\sin 2\varphi\left\{y_{t10}^{12}\frac{1}{\varepsilon s}\right.\\&\cdot\left[1-\frac{1}{w}-\frac{2c}{\varepsilon s^2}(1-\varepsilon c-w)\right]+y_{t11}^{12}\left[\left(\frac{c}{\varepsilon^2s^3}\right.\right.\\&\left.+\frac{1}{2\varepsilon^2s^3w}\left(\varepsilon\left(1+c^2\right)-2c+\frac{\varepsilon s^2}{w^2}(\varepsilon c-1)\right)\right)\\&+\frac{\varepsilon s}{2}\left(\frac{1}{w^3}+\frac{2\left(1+w^{-1}\right)+w^{-3}\varepsilon c}{1+w-\varepsilon c}+\frac{\varepsilon c\left(1+w^{-1}\right)^2}{(1+w-\varepsilon c)^2}\right)\\&\left.\left.-\frac{1}{\varepsilon s}\left(1-\frac{1}{w}+\frac{\varepsilon(\varepsilon-c)}{w^3}+\frac{2c}{s^2}\left(c+\frac{\varepsilon-c}{w}\right)\right)\right]\right\}\end{aligned} \tag{4d}$$

$$\begin{aligned}u_\varphi^{t,12}(a,\theta,\varphi)&-2\cos 2\varphi y_{t10}^{12}\frac{1}{\varepsilon s}\left[\frac{\varepsilon s^2}{w^3}\right.\\&\left.+\frac{2}{\varepsilon s^2}(2-s^2)(1-\varepsilon c-w)-2c\left(1-\frac{1}{w}\right)\right]\end{aligned} \tag{4e}$$

$$\begin{aligned}u_r^{32}(a,\theta,\varphi)&=2\sin\varphi\\&\cdot\left[y_{240}^{23}\frac{s\left(1+\varepsilon c-2\varepsilon^2\right)}{\varepsilon w^5}+y_{241}^{23}\frac{s}{\varepsilon w^3}\right]\end{aligned} \tag{4f}$$

$$u_{\theta}^{s,32}(a,\theta,\varphi)=2\sin\varphi\cdot\left\{y_{340}^{23}\frac{1}{\varepsilon w^{5}}\left[c\left(1+\varepsilon^{2}\right)+\varepsilon\left(c^{2}-3\right)\right]+y_{341}^{23}\frac{1}{\varepsilon w^{2}(1-\varepsilon c+w)^{2}}\cdot\left[\left(cw(1+w)+\varepsilon s^{2}\right)(1-\varepsilon c+w)-\varepsilon s^{2}(1+w)\left(w+\varepsilon+\frac{1}{w}(1-\varepsilon c+w)\right)\right]+y_{342}^{23}\frac{1}{\varepsilon^{3}}\left[\varepsilon\left(\frac{1}{w}-1\right)-\frac{c}{s^{2}}(\varepsilon c-1+w)\right]\right\} \quad (4g)$$

$$u_{\varphi}^{s,32}(a,\theta,\varphi)=2\cos\varphi\left[y_{340}^{23}\frac{1}{\varepsilon w^{3}}+y_{341}^{23}\cdot\frac{1+w}{\varepsilon w(1-\varepsilon c+w)}+y_{342}^{23}\frac{1}{\varepsilon^{3}s^{2}}(\varepsilon c-1+w)\right] \quad (4h)$$

$$u_{\theta}^{t,32}(a,\theta,\varphi)=-2\sin\varphi\left\{y_{t20}^{23}\frac{1}{\varepsilon^{2}s^{2}}\cdot\left(1-\frac{1-\varepsilon c}{w}\right)+y_{t21}^{23}\left[\frac{-1}{\varepsilon^{2}s^{2}}(1-\varepsilon c-w)\right]\right\} \quad (4i)$$

$$u_{\varphi}^{t,32}(a,\theta,\varphi)=-2\cos\varphi\left\{y_{t10}^{23}\frac{-1}{\varepsilon^{2}s^{2}}\cdot\left[c+\frac{1}{w}\left(\varepsilon s^{2}-(1-\varepsilon c)\left(c+\frac{\varepsilon s^{2}}{w^{2}}\right)\right)\right]+y_{t11}^{23}\left[\frac{c}{\varepsilon^{2}s^{2}}(1-\varepsilon c-w)-\frac{1}{\varepsilon}\left(1-\frac{1}{w}\right)\right]\right\} \quad (4j)$$

$$u_{r}^{22,0}(a,\theta,\varphi)=\left(y_{220}^{22}+y_{230}^{22}\right)\frac{1}{w^{5}}\left[c(1-\varepsilon^{2})+\varepsilon(c^{2}-2+\varepsilon^{2})\right]+\left(y_{210}^{22}+y_{221}^{22}+y_{230}^{22}+y_{231}^{22}\right)\frac{1}{w^{3}}(c-\varepsilon)+\left(y_{211}^{22}+y_{222}^{22}+y_{231}^{22}+y_{232}^{22}\right)\frac{1}{\varepsilon w}+\left(y_{212}^{22}+y_{232}^{22}\right)\frac{1}{\varepsilon}(\ln 2-\ln(1-\varepsilon c+w)) \quad (4k)$$

$$u_{\theta}^{s,22,0}(a,\theta,\varphi)=\left(y_{320}^{22}+y_{330}^{22}\right)\frac{-s}{w^{5}}\left(1+\varepsilon c-2\varepsilon^{2}\right)+\left(y_{310}^{22}+y_{321}^{22}+y_{330}^{22}+y_{331}^{22}\right)\frac{-s}{w^{3}}+\left(y_{311}^{22}+y_{331}^{22}+y_{332}^{22}\right)\frac{-s(1+w)}{w(1-\varepsilon c+w)} \quad (4l)$$

$$u_{\varphi}^{s,22,0}(a,\theta,\varphi)=u_{\theta}^{t,22,0}(a,\theta,\varphi)=u_{\varphi}^{t,22,0}(a,\theta,\varphi)=0 \quad (4m)$$

$$u_{r}^{33}(a,\theta,\varphi)=y_{220}^{33}\frac{1}{\varepsilon w^{5}}\cdot\left[c(1-\varepsilon^{2})+\varepsilon(c^{2}-2+\varepsilon^{2})\right]+y_{221}^{33}\frac{1}{\varepsilon w^{3}}(c-\varepsilon)+y_{222}^{33}\frac{1}{\varepsilon^{2}w} \quad (4n)$$

$$u_{\theta}^{s,33}(a,\theta,\varphi)=y_{320}^{33}\frac{-s}{\varepsilon w^{5}}\left(1+\varepsilon c-2\varepsilon^{2}\right)+y_{321}^{33}\frac{-s}{\varepsilon w^{3}} \quad (4o)$$

$$u_{\varphi}^{s,33}(a,\theta,\varphi)=u_{\theta}^{t,33}(a,\theta,\varphi)=u_{\varphi}^{t,33}(a,\theta,\varphi)=0 \quad (4p)$$

where the symbols y_{klm}^{ij} are constants and to be determined by the parameters of an earth model (Sun et al., 1996). Notice that the case of $u^{22}(a,\theta,\varphi)$ is slightly special because the source function contains two parts with m=0 and m=2, respectively. Formulas (4k-m) give solutions for the case of m=0, while solutions of the case m=2 can be derived from $u^{12}(a,\theta,\varphi)$ according to the relation of the source functions. Expressions (4a-p) show the main results of this research. They can be used to calculate displacement components excited by four types of sources buried in a spherically symmetric earth model. In combination, these components allow calculation of a displacement field excited by an arbitrary seismic source. It is seen from the expressions in Eqs. (4a-p) that no numerical integration or summation is necessary. Once an earth model and source parameters are provided, excited displacements can be calculated analytically. These are merits of asymptotic solutions. Note that the co-seismic displacement for a limited fault can be obtained easily by integration of the point dislocation over the fault plane.

4. Numerical Test for Homogenous Earth Model

The theory proposed in this research is valid for an arbitrary spherically symmetric earth model provided the parameters of the earth model could be continuously and smoothly described since derivatives of earth surface parameters are needed. In practice, however, there is no such earth model available. For example, earth model 1066A is given in a layered structure with 160 layers, where the top layer of 11km is homogenous. In this case, derivatives vanish and the layered inner structure cannot be fully reflected. Therefore, as an example, here we consider a homogenous sphere with identical parameters to the 1066A top layer. Using the homogenous sphere and the above theory, we perform numerical calculation to check the validity of the theory for a practical application. We consider displacements calculated for 32 km depth, we obtain corresponding displacements for the vertical strike-slip source (as an example) in Fig. 2 using numerical summations (Sun et al., 1996) and

the above analytical formulas (4), respectively. Results are given for $\sin 2\varphi = 1$, i.e., $\varphi = 45°$. The x-axis stands for the epicentral distance up to three degrees; the y-axis represents magnitudes of displacement in meters. The solid line shows asymptotic displacements, while the dashed line shows integrated displacements. Results are seen to agree very well; discrepancies are too small to be identified in these figures because the two curves overlap each other.

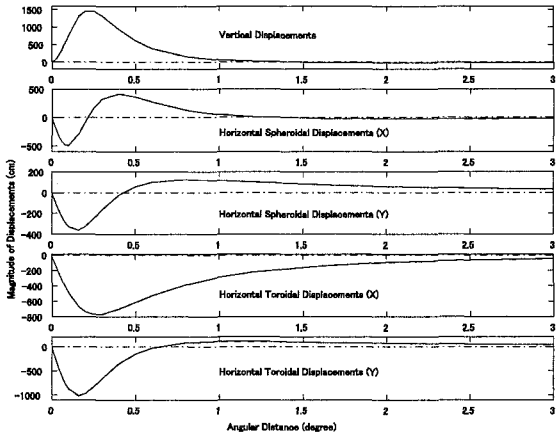


Fig. 2. Comparisons of displacements calculated by asymptotic and integrated methods for a vertical strike-slip dislocation at 32km depth in a homogeneous sphere.

The discrepancy between the integrated and asymptotic results can be observed by plotting their distribution in far field. For this purpose, we plot in Figs. 3 and 4 the vertical and horizontal displacements, respectively, caused by a vertical strike-slip source at a depth of 32 km. Note that the horizontal displacement is composed of both spheroidal and toroidal components and at direction of $\varphi = 45°$. The dashed line stands for the integrated results; the solid line represents the asymptotic results. These results are plotted for the entire global surface with different epicentral distance segments, i.e., 0°—3°, 3°—10°, 10°—30°, 30°—90° and 90°—180°. For comparison, the flat-earth result is also plotted in the same figure (the dashdot line). Comparison between the solid and dashed lines shows that the asymptotic results agree well with the integrated one. Only a slight discrepancy appears after 30° for the vertical displacement, and even no big difference until 90° for the horizontal displacement. This means that the asymptotic theory for a vertical strike-slip fault basically applies to the whole earth. If higher order terms in the asymptotic solutions are included, the asymptotic theory is expected to approximate the exact theory even better.

On the other hand, Figs. 3 and 4 also shows that the asymptotic results approximate the exact results much better than those by the flat-earth results, since the discrepancy between the flat-earth and exact results becomes larger and larger as the epicentral distance increases. For the vertical displacement, the flat-earth gives wrong results (with an opposite sign) outside about 60°. This is natural due to the lack of the sphericity effect in the flat-earth model. This proves our previous conclusion (Sun and Okubo, 2002), i.e., there exists a big difference between displacements obtained for a half-space and a homogenous earth, and the deeper the source depth, the larger the spherical effect. Since the above asymptotic displacements include curvature effects (and also self-gravity effect), the asymptotic theory for a homogeneous sphere is theoretically more reasonable than that for a half space model (Okada, 1985). Also, the applicability of this theory for any epicentral distance is superior to Okada (1985), which is applicable only for the near field. Furthermore, the asymptotic theory is mathematically simple, so that the practical application becomes as simple as the theory for a half space.

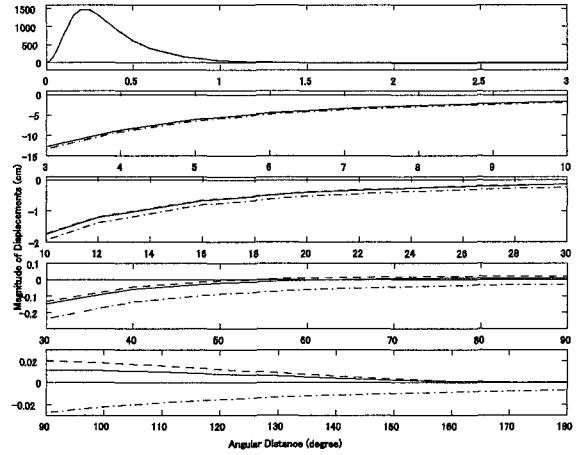


Fig. 3. Comparison between the vertical displacements caused by a vertical strike-slip source at a depth of 32 km, calculated by three dislocation theories: the exact dislocation theory for a homogeneous sphere, the asymptotic dislocation theory for the same model, and the dislocation theory for a flat-earth model.

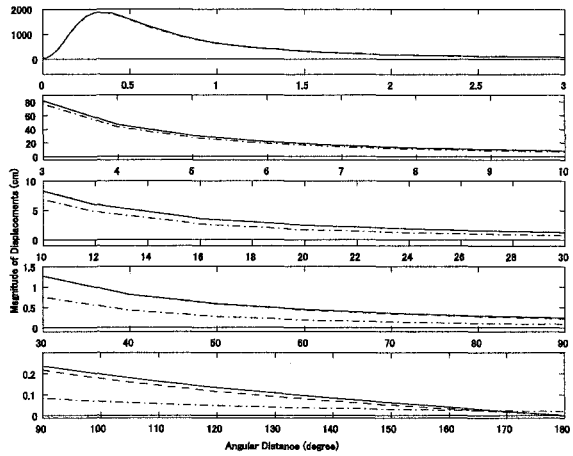


Fig. 4. Comparison of the horizontal displacements caused by a vertical strike-slip source at a depth of 32 km, calculated by three dislocation theories.

5. Conclusions

This research presents a new theory for calculating asymptotic displacements excited by a point dislocation in a spherical symmetric earth model as an approximation of the dislocation theory for a spherical earth model (Sun et al., 1996); it uses the reciprocity theorem (Okubo, 1993) and asymptotic solutions of an elastic earth to tidal, press, and shear forces (Okubo, 1988). Final conclusions and remarks are: 1) The asymptotic theory is given in an analytical form, with no numerical integration and summation, so that it can be used as easily as the theory for a half space model, such as Okada (1985). 2) This asymptotic theory principally applies for a spherically symmetric earth model provided the parameters of the model can be expressed in continuous functions of the earth's radius. 3) The numerical results in this research are calculated for a homogenous earth model. They approximate the exact results of a spherical earth much better that those calculated by a flat earth theory for any epicentral distance. 4) Both theoretical and numerical investigations indicate that this theory is valid for an epicentral distance of at least 20° for all types of seismic sources, with a relative error less than 1%. 5) The asymptotic solutions presented in this study can also be used for the first approximation of an exact solution of a real spherical earth model. It means that a more realistic result can be obtained by adding a small correction to the asymptotic one.

Acknowledgements. Useful comments by two anonymous reviewers are gratefully acknowledged. This research was financially supported by JSPS research grants (C13640420) and "Basic design and feasibility studies for future missions for monitoring Earth's environment".

References

Ben-Menahem, A. & Singh S. J., 1968. Eigenvector expansions of Green's dyads with applications to geophysical theory, Geophys. J. R. Astron. Soc., 16, 417-452.

Gilbert F. & Dziewonski A.M., 1975. An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra, Phil. Trans. R. Soc London A, 278, 187-269.

Maruyama, T., 1964. Statical elastic dislocations in an infinite and semi-infinite medium, Bull. Earthquake Res. Inst. Univ. Tokyo, 42, 289-368.

Okada, Y., 1985. Surface deformation due to shear and tensile faults in a half-space, Bull. Seism. Soc. Am., 75, 1135-1154.

Okubo, S., 1988. Asymptotic solutions to the static deformation of the Earth, 1, Spheroidal mode, Geophys. J. Int., 92, 39-51.

Okubo, S., 1991. Potential and gravity changes raised by point dislocations, Geophys. J. Int., 105, 573-586.

Okubo, S., 1992. Potential gravity changes due to shear and tensile faults, J. Geophys. Res., 97, 7137-7144.

Okubo, S., 1993. Reciprocity theorem to compute the static deformation due to a point dislocation buried in a spherically symmetric Earth, Geophys. J. Int., 115, 921-928.

Piersanti, A., G. Spada, R. Sabadini, & Bonafede M., 1995. Global post-seismic deformation, Geophys. J. Int., 120, 544-566.

Piersanti, A., G. Spada & Sabadini R., 1997. Global postseismic rebound of a viscoelastic Earth: Theory for finite faults and application to the 1964 Alaska earthquake, J. Geophys. Res. 102, 477-492.

Saito, M., 1967. Excitation of free oscillations and surface waves by a point source in a vertically heterogeneous Earth, J. Geophys. Res., 72, 3689-3699.

Sabadini, R., Piersanti A. & Spada G., 1995. Toroidal/poloidal partitioning of global Post-seismic deformation, Geophys. Res. Lett., 21, 985-988.

Smylie, D.S., & Mansinha L., 1971. The elasticity theory of dislocation in real Earth models and changes in the rotation of the Earth, Geophys. J. R. Astron. Soc., 23, 329-354.

Steketee, J. A., 1958. On Volterra's dislocations in a semi-infinite elastic medium, Can. J. Phys., 36, 192-205.

Sun W. & Okubo, S., 1993. Surface potential and gravity changes due to internal dislocations in a spherical earth – I. Theory for a point dislocation, Geophys. J. Int., 114, 569-592.

Sun, W., Okubo S. & Vanicek, P., 1996. Global displacement caused by dislocations in a realistic earth model, J. Geophys. Res., 101, 8561-8577.

Sun, W. & Okubo S., 2002. Effects of the Earth's Spherical Curvature and Radial Heterogeneity in Dislocation Studies – For a Point Dislocation, Geophys. R.L., in press.

The Use of Energy Balance Relations for Validation of Gravity Field Models and Orbit Determination Results

K.H. Ilk, A. Löcher
Institute of Theoretical Geodesy, University of Bonn, Nussallee 17, D-53115 Bonn, Germany

Abstract. There is a need for a proper validation procedure of gravity field solutions, especially for those high precise ones which are derived from the dedicated gravity field missions as CHAMP, GRACE and in future GOCE. In this paper the balance equations for the energy and for the energy exchange of a satellite orbiting around the Earth are proposed as analysis tool to validate gravity field solutions together with orbits derived with precise reduced dynamic or kinematic orbit determination procedures. The theoretical foundation of the analysis tool is presented as well as simulation results and applications to real orbits of a GPS satellite and of CHAMP based on various gravity field solutions. It is shown that the validation procedure can be used to detect deficiencies in the orbit modelling and in gravity field recovery results.

Keywords. Gravity field validation – Orbit determination validation – Energy balance relations – Jacobi-Integral

1 Introduction

The satellite missions CHAMP, GRACE and GOCE will provide unprecedented views of the Earth's gravity field and its changes with time. For many applications in the Earth sciences the new satellite solutions shall not be combined with available gravity field information, to provide consistent and unbiased satellite-only results of the gravity field. Therefore, a natural control of the recovery results by alternative data sets does not exist. Evidently, there is an urgent need for a proper validation tool for gravity field solutions of high precision.
The validation and verification procedures based on comparisons with existing models in well-determined regions or based on orbit predictions may not fulfill the expectations for a rigorous validation: Available high-frequency solutions derived from satellite perturbations and altimetry and combined with terrestrial gravity field information are not consistent because of different datum references in space and time. Furthermore, the computation of gravity field solutions based on satellite techniques require the regularization of the large and ill-posed normal equations, which produces a bias in the results and can mislead the interpretation of the gravity field. On the other hand, comparisons on the basis of orbit computations by using a gravity field model are susceptible to numerical errors and the orbit residuals caused by incorrect force models may hide the real causes of deficiencies.
The energy exchange relations of the satellites' motion within the gravity field of the Earth seem to be useful for the validation and the consistency check of gravity field models and orbit determination results when applied in the sense of a forward computation process. One of the advantages is the more or less point wise application along the orbit, avoiding any error accumulation or instability effects. Besides the validation tasks the energy relations can represent the basis for a subsequent combination with alternative and additional gravity field information and may be used to calibrate measurement procedures.
The application of the energy integral for problems of satellite geodesy has been proposed since its very beginning (e.g., O'Keefe, 1959, Bjerhammar, 1967, Reigber, 1969). But the applications did not lead to convincing results because of the type of observations and the poor coverage of the satellite orbits with observations available at that time. The situation changed with the new type of homogeneous and dense data distributions as recently demonstrated e.g. by Jekeli (1999).

2 Gibbs function and Gibbs fundamental form

A physical system can be characterized by its Gibbs' function, a function which represents the total energy of the system (Falk and Ruppel, 1973). The Gibbs function is formulated in terms of "extensive" quantities X_i, $E(X_1,....,X_n)$. In those cases where all quantities are well-known, the physical system is completely described, as well as the processes, in which the system can exchange dynamical quantities such as energy, linear momentum or angular momentum. The exchange of energy then occurs in various energy forms. Any energy form is defined by a pair of conjugate variables, the "intensive" and the "extensive" quantities. The exchange of energy is expressed by the change of the extensive variable. The product of an intensive quantity and the differential of an extensive variable results in the energy form. The change dE of the total energy E of the system is formulated by

Gibbs fundamental form:

$$dE = \sum_{j=1}^{n} \xi_j \, dX \,. \qquad (1)$$

The intensive variables are derived as partial derivatives of Gibbs function with respect to the conjugate extensive variables

$$\xi_j = \frac{\partial E(X_1, \ldots\ldots, X_n)}{\partial X_j} \,. \qquad (2)$$

The Gibbs function can be identified with the Hamiltonian function and the generalized impulses and coordinates with the extensive variables, while the time derivatives of the latter quantities correspond to the intensive variables. If we consider the simple case of a satellite, orbiting around a rigid Earth, and if we take into account only mechanical energy forms then the Gibbs function reads (Ilk, 2002, 2003),

$$\hat{E} = \frac{1}{2}\frac{\mathbf{p}_E^2}{M} + \frac{1}{2}\frac{\mathbf{P}^2}{M} + \frac{1}{2}\mathbf{L}_E \cdot \mathbf{T}_E^{-1} \cdot \mathbf{L}_E + \hat{V}(\mathbf{R}, \boldsymbol{\Omega}) = const \,, \quad (3)$$

with the linear momentum $\mathbf{p}_E$ of the Earth, the angular momentum $\mathbf{L}_E$ of the Earth and $\mathbf{T}_E$ its inertia tensor. $\mathbf{P}$ is the linear momentum of the satellite referred to the geo center and M its mass. $V(\mathbf{R}, \boldsymbol{\Omega})$ is the potential energy of the satellite with its position $\mathbf{R}$ and the rotation vector $\boldsymbol{\Omega}$ of the Earth. Linear and angular momentum of the Earth can be considered as constants and included in the energy constant, so that it follows the relation,

$$\hat{E} = \frac{1}{2}\frac{\mathbf{P}^2}{M} + \hat{V}(\mathbf{R}, \boldsymbol{\Omega}) = const \,, \qquad (4)$$

which represents the well known energy balance for this simple case formulated with respect to an inertial system. In case of a rotating reference system an inertia field is involved. In case of a constant rotation expressed by the rotation vector $\boldsymbol{\Omega}$ the energy balance reads

$$\hat{E}' = \frac{1}{2}\frac{\mathbf{P}'^2}{M} - \mathbf{P}' \cdot (\boldsymbol{\Omega} \times \mathbf{R}') + \hat{V}'(\mathbf{R}') = const \,, \qquad (5)$$

where quantities referring to the rotating reference system are marked by an inverted comma. The potential energy $V'(\mathbf{R}')$ is related to the gravitational potential $V'(\mathbf{R}')$ of the Earth by the relation

$$\hat{V}'(\mathbf{R}') = -MV'(\mathbf{R}') \,. \qquad (6)$$

The velocity of the satellite can be derived from the Gibbs function by differentiation with respect to the linear momentum ($\mathbf{P}' = \mathbf{P}$) due to (2), so that it holds:

$$\dot{\mathbf{R}}' = \frac{\partial \hat{E}'}{\partial \mathbf{P}'} = \frac{\mathbf{P}'}{M} - \boldsymbol{\Omega} \times \mathbf{R}' = \dot{\mathbf{R}} - \boldsymbol{\Omega} \times \mathbf{R}' \,, \qquad (7)$$

where $\dot{\mathbf{R}}$ is the velocity referred to the inertial system. Inserting this relation into formula (5) results in an energy balance relation, the so-called Jacobi integral,

$$\hat{E}' = \frac{M}{2}\dot{\mathbf{R}}'^2 - \frac{M}{2}(\boldsymbol{\Omega} \times \mathbf{R}')^2 - MV'(\mathbf{R}') = const \,, \qquad (8)$$

and divided by the constant satellite mass M,

$$E' = \frac{1}{2}\dot{\mathbf{R}}'^2 - \frac{1}{2}(\boldsymbol{\Omega} \times \mathbf{R}')^2 - V'(\mathbf{R}') = const \,. \qquad (9)$$

The rotational term represents the centrifugal potential und can be added to the gravitational potential to result in the gravity potential

$$W(\mathbf{R}') = \frac{1}{2}(\boldsymbol{\Omega} \times \mathbf{R}')^2 + V'(\mathbf{R}') \,, \qquad (10)$$

so that the (specific) energy balance can be written in the form

$$E' = \frac{1}{2}\dot{\mathbf{R}}'^2 - W(\mathbf{R}') = const \,. \qquad (11)$$

3 Energy Integral for Satellite Motion

Unfortunately, the Gibbs function is not accessible in an easy way if the motion of an artificial satellite within the gravity field of the Earth has to be considered under real conditions. Because of the various rotational constituents of the Earth, such as daily rotation, polar motion and precession-nutation, the rotation vector $\boldsymbol{\Omega}$ is not constant. Furthermore, in the case of the satellite mission CHAMP but also GRACE, accelerations caused by surface forces such as air drag and solar pressure acting on the satellites are measured and it is not easy to formulate the corresponding functional constituents in the Gibbs function. If the surface accelerations are not observable in-situ then models for these kinds of forces can be used; again the corresponding energy functions are not available directly. Additional problems occur if the physical system has to include also Sun and Moon and the various time dependent and therefore non-conservative tidal constituents acting on the motion of the satellite. In this case and in case of additional non-conservative forces $\mathbf{K}'$ it is preferable to compute the energy contributions by the work $A|_{t_0}^{t}$ performed by the various forces along the satellite orbit within the time period $[t_0, t]$ (Löcher, unpublished manuscript):

$$\hat{E}'_K|_{t_0}^{t} = -A'|_{t_0}^{t} = -\int_{t_0}^{t} \mathbf{K}' \cdot d\mathbf{R}' = -\int_{t_0}^{t} \mathbf{K}' \cdot \dot{\mathbf{R}}' dt \,. \qquad (12)$$

Inserting the velocity according to equation (7) this formula reads

$$\hat{E}'_K|_{t_0}^{t} = -\int_{t_0}^{t} \mathbf{K}' \cdot d\mathbf{R}' = -\int_{t_0}^{t} \mathbf{K}' \cdot \left(\frac{\mathbf{P}'}{M} - \boldsymbol{\Omega} \times \mathbf{R}'\right) dt \,. \qquad (13)$$

The contribution of the centrifugal acceleration $\mathbf{Z}'$ to the energy balance is already included in the energy balance and corresponds to the rotational term in the Jacobi integral in form of formula (9). There is no contribution of the Coriolis acceleration $\mathbf{C}'$. In case of temporal variations of the rotation vector $\boldsymbol{\Omega}$ the contribution of the Eulerian acceleration,

$$\mathbf{E}' = -M\dot{\boldsymbol{\Omega}}' \times \mathbf{R}' = -M\dot{\boldsymbol{\Omega}} \times \mathbf{R} \,, \qquad (14)$$

to the energy balance has to be taken into account as

well,

$$\hat{E}'_E|'_{t_0} = -\int_{t_0}^{t} \mathbf{E}' \cdot \dot{\mathbf{R}}\, dt = M \int_{t_0}^{t} \dot{\boldsymbol{\Omega}} \cdot \left(\mathbf{R} \times \dot{\mathbf{R}}\right) dt \,. \qquad (15)$$

This term is not covered by the Jacobi integral (8) and there is no scalar potential for the acceleration function $\mathbf{E}'$. Therefore, the energy contribution has to be computed numerically point wise along the orbit. The same holds also for the surface forces and the various tidal forces acting on the satellite. For example, the tidal forces exerted by the Moon or the Sun (mass M_i, $i = M,S$) onto the satellite's motion read with sufficient approximation,

$$\mathbf{G}'_{(Sat,E)i} = -GMM_i \left(\frac{\mathbf{R}' - \mathbf{R}'_i}{\left|\mathbf{R}' - \mathbf{R}'_i\right|^3} + \frac{\mathbf{R}'_i}{\left|\mathbf{R}'_i\right|^3} \right), \qquad (16)$$

resulting in an energy contribution according to (12),

$$\hat{E}'_i|'_{t_0} = -\int_{t_0}^{t} \mathbf{G}'_{(Sat,E)i} \cdot \left(\dot{\mathbf{R}} - \boldsymbol{\Omega} \times \mathbf{R}'\right) dt \,. \qquad (17)$$

The terms (15) and (17) are not covered by the Jacobi integral as well; therefore an extended Jacobi integral will be introduced,

$$J := E'(t) + E'_E|'_{t_0} + E'_M|'_{t_0} + E'_S|'_{t_0} = const \,. \qquad (18)$$

with E' according to formulae (9) or (11) and the specific energy quantities,

$$E'_i|'_{t_0} := \hat{E}'_i|'_{t_0} / M \,. \qquad (19)$$

This relation shall be applied in the next paragraph where a simulation scenario has been constructed.

4 Simulation scenario

To get an impression of the size of the various energy contributions to the energy balance a simplified model consisting of an artificial satellite orbiting around the Earth under the gravitational influence of the Sun and the Moon has been simulated. Sun, Moon and the satellite are modeled as point masses. The motions of all bodies involved in the simulation scenario have been described with reference to an Earth fixed reference system. The equation of motion for the Sun with mass M_S reads

$$M_S \frac{d\dot{\mathbf{R}}'_S}{dt} = \mathbf{K}'_S + \mathbf{Z}'_S + \mathbf{C}'_S + \mathbf{E}'_S \,, \qquad (20)$$

with the gravitational forces between Earth and Sun

$$\mathbf{K}'_S = -GM_S M_E \frac{\mathbf{R}'_S}{\left|\mathbf{R}'_S\right|^3} \,, \qquad (21)$$

and the inertia accelerations $\mathbf{Z}'_S, \mathbf{C}'_S, \mathbf{E}'_S$. Analogously, the equation of motion for the Moon reads

$$M_M \frac{d\dot{\mathbf{R}}'_M}{dt} = \mathbf{K}'_M + \mathbf{G}'_{(M,E)S} + \mathbf{Z}'_M + \mathbf{C}'_M + \mathbf{E}'_M \,, \qquad (22)$$

with the gravitational force between Earth and Moon $\mathbf{K}'_M$, the tidal force exerted by the Sun onto the Moon's motion, $\mathbf{G}'_{(M,E)S}$, and the inertia accelerations $\mathbf{Z}'_M, \mathbf{C}'_M, \mathbf{E}'_M$. Finally, the equation of motion for the satellite reads

$$M \frac{d\dot{\mathbf{R}}'}{dt} = \mathbf{K}'_{Sat} + \mathbf{G}'_{(Sat,E)M} + \mathbf{G}'_{(Sat,E)S} + \mathbf{Z}' + \mathbf{C}' + \mathbf{E}' \,, \qquad (23)$$

with the gravitational force between satellite and Earth $\mathbf{K}'_{Sat}$, expressed by

$$\mathbf{K}'_{Sat} = \nabla V'(\mathbf{R}') \,, \qquad (24)$$

with the gravitational potential of the Earth V'.

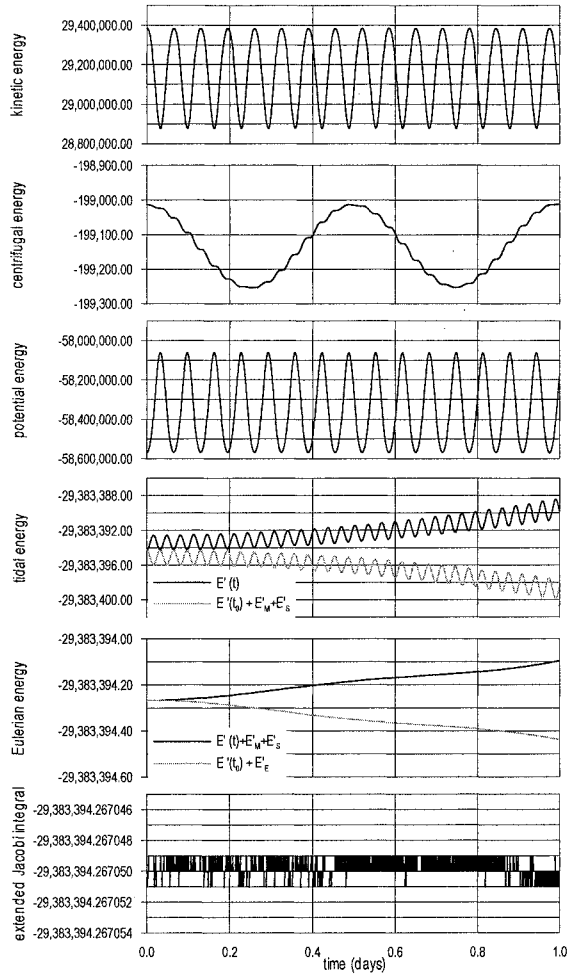


Fig. 1: Energy contributions (m / s) to the extended Jacobi integral for a simulated one-day arc of CHAMP (according to equations (5) and (18)).

Euler's equation of motion for the rotation of a rigid body has been applied to simulate the rotation of the Earth and, fixed to it, the rotation of the terrestrial reference frame:

$$\begin{aligned} A\dot{\Omega}_x + (C-B)\Omega_y\Omega_z &= M'_x, \\ B\dot{\Omega}_x + (A-C)\Omega_x\Omega_z &= M'_y, \\ C\dot{\Omega}_z &= M'_z, \end{aligned} \qquad (25)$$

with the principal moments of inertia A,B,C and the coordinates of the torque M_x, M_y, M_z exerted onto the Earth by Sun and Moon. The coupled system of ordinary differential equations (20), (22), (23) and (25) has been integrated numerically with a Runge-

Kutta integrator of 4th order with a step size of 1*s* and for a period of one day. The result of the computation builds the basis for the computation of the various energy constituents according to formula (18). Fig. 1 shows the results for the extended Jacobi integral with its various energy contributions. The graphs demonstrate the temporal evolution of the energy contributions during one day and the corresponding energy exchange behavior. At the same time they proof the basic procedure for the computation of the energy balance along a satellite orbit.

5 Application to observed orbits

In this paragraph the procedure shall be applied to the computation of the extended Jacobi integral along an arc of the so-called Post-processed Science Orbits (PSO) of a GPS satellite and of an arc of the PSO of CHAMP. A special feature of the CHAMP mission is the fact that all surface forces are measured by a three-dimensional accelerometer. These data sets as well as observations and orbits at different production levels are provided by the CHAMP Information System and Data Centre (ISDC). From the surface forces acting on the GPS satellites only the solar radiation pressure has to be modeled; the air drag is not relevant at the altitude of these satellites. It turned out that also the Earth and ocean tides have no significant influence on the satellite orbits of the GPS.

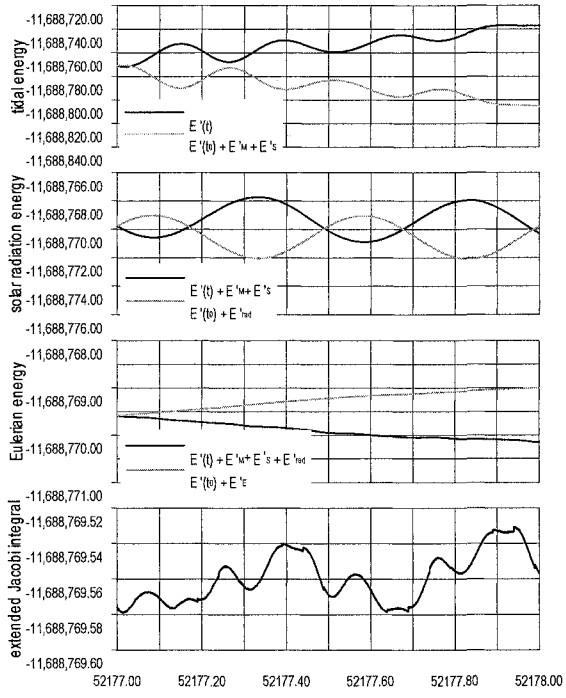


Fig. 2: Energy contributions (m / s) to the extended Jacobi integral for a one-day PSO arc of GPS 10.

Therefore, besides the kinetic energy, the centrifugal energy and the potential energy according to formula (5), only the energy contributions from the direct tidal forces ($E'_M|^t_{t_0}$, $E'_S|^t_{t_0}$), from the Eulerian acceleration ($E'_E|^t_{t_0}$), and from the solar pressure ($E'_{rad}|^t_{t_0}$) have to be considered for the extended Jacobi integral according to formula (18):

$$J_{GPS} := E'(t) + E'_M|^t_{t_0} + E'_S|^t_{t_0} + E'_E|^t_{t_0} + E'_{rad}|^t_{t_0} = const. \quad (26)$$

The extended Jacobi integral for the CHAMP satellite orbit includes instead of the solar pressure the energy contributions by the measured accelerations ($E'_{akz}|^t_{t_0}$) as well as the contributions from the Earth tides ($E'_{Earth\,Tides}|^t_{t_0}$) and the ocean tides ($E'_{Ocean\,Tides}|^t_{t_0}$) and all other energy forms as used for the GPS satellite resulting in the extended Jacobi integral:

$$J_{CHAMP} := E'(t) + E'_{acc}|^t_{t_0} + E'_M|^t_{t_0} + E'_S|^t_{t_0} + E'_E|^t_{t_0} + \\ + E'_{Earth\,Tides}|^t_{t_0} + E'_{Ocean\,Tides}|^t_{t_0} = const. \quad (27)$$

The orbits of the CHAMP data base are referred to a terrestrial reference frame due to Lühr et al., 2001. The transformation between the terrestrial and celestial reference frames followed the IERS conventions published by McCarthy, 1996. The computation of the tidal energy requires the ephemeris of Moon and Sun. For this task the numerical ephemeris DE405 of the Jet Propulsion Laboratory (JPL) have been used together with the software to perform the processing of these data. The computation of the solar radiation pressure has been based on the CODE-RPR model of the Center of Orbit Determination in Europe (CODE) (Springer et al., 1998) by applying the shadow function according to Montenbruck and Gill (2000). The accelerometer measurements, needed to determine the surface force energy for the CHAMP orbit, have been processed according to the rules of the CHAMP data formats supplied by Förste et al., 2002. Finally, the energy contributions caused by the Earth tides as well as by the ocean tides have been based on the models as published by McCarthy, 1996. With these extensive computations the extended Jacobi integral of a one-day arc (25.9.2001) of the PSO of GPS 10 has been determined (Löcher, unpublished manuscript). The results for the GPS satellite are shown in Fig. 2. The deviations from a Jacobi constant show still systematic effects within a band of 0,04 m^2/s^2 which can be translated into a variation of the orbit positions of approximately 7cm; this corresponds to the 3cm which are given in the CHAMP data base. Improvements can be expected with more precise solar pressure models as well as with improved techniques for the derivation of the velocities of the GPS satellites. The contributions to the extended Jacobi integral in case of a two-revolution arc (1.12.2001) of the PSO of CHAMP are shown in Fig. 3 (due to (5) and (26)). As gravity field model the recent satellite-only gravity field solution EIGEN-2 has been used for this test (Reig-

ber et al., 2003). The secular trend at the bottom graph could be caused by an insufficient calibration of the accelerometer.

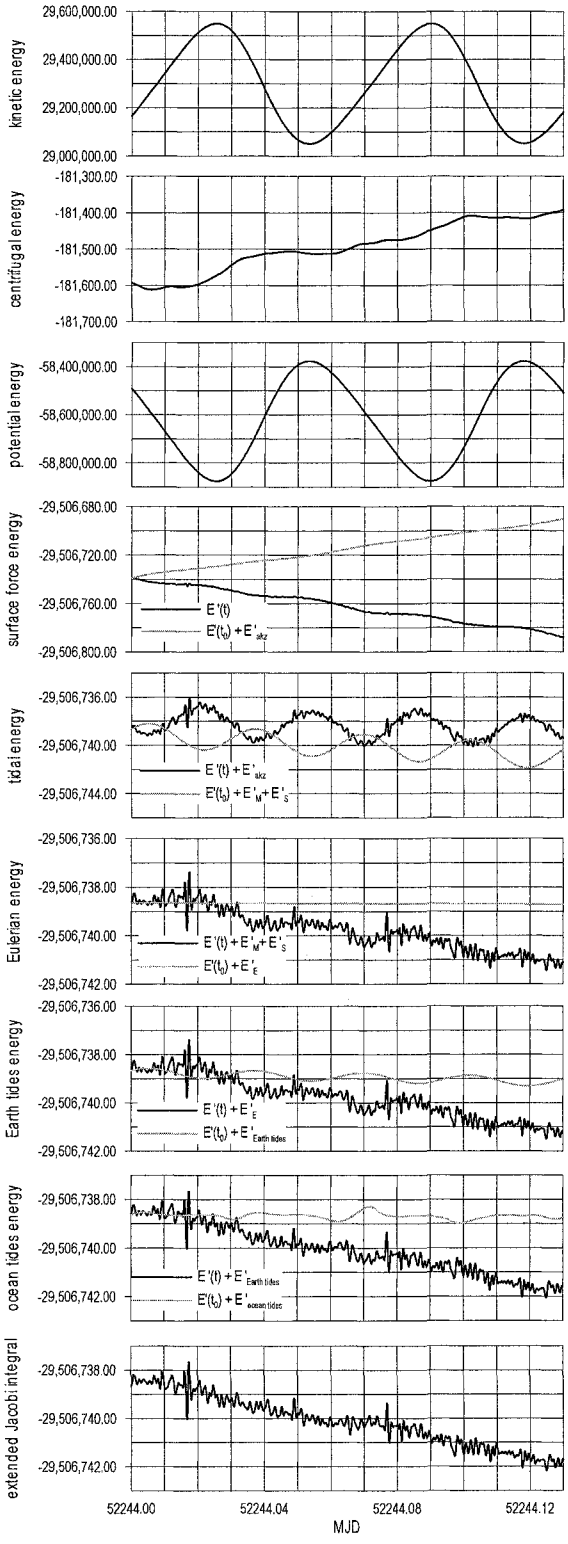


Fig. 3: Energy contributions (m / s) to the extended Jacobi integral for a two-revolution PSO arc of CHAMP.

If this linear trend has been removed then a deviation from a constant at a size of 0,26 m^2/s^2 remains, which corresponds to approximately 3cm in the positions. While in case of the GPS satellite the main error source may come from the orbit models and the orbit determination technique in case of CHAMP the gravity field will have an important influence on the energy balance as well as the various force models introduced. The influence of various gravity fields on the Jacobi constant is demonstrated in Fig. 4. While the gravity field model OSU91 but also EGM96 show significant deviations from a constant, the results based on the CHAMP model EIGEN-1S (Reigber et al., 2002) and the recent model EIGEN-2 (Reigber et al., 2003) show clear improvements after removal of the secular trend. But also in case of EIGEN-2 some features of the residuals could be caused by an insufficient modeling of the orbits. The reason might be regional rough gravity field features which can not be modeled by a global gravity field representation with spherical harmonics. A focusing regional gravity field recovery by space localizing base functions as harmonic spline functions show significant improvements as demonstrated in Ilk et al., 2003.

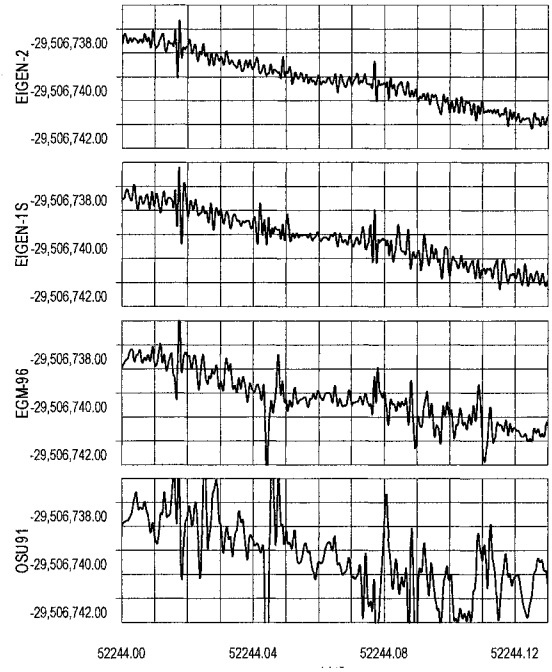


Fig. 4: Extended Jacobi integral (m / s) for a two-revolution PSO arc of CHAMP in case of various gravity field models.

The results achieved so far reflect the high accuracy of the PSO determination as well as of the gravity model EIGEN-2. In case of the less precise Rapid Science Orbits (RSO) the graphs are significantly rougher as a comparison of PSO-RSO for two revolutions shows (Fig. 5). The error for the RSO

(grey) is twice as large as those for the PSO (black). A closer look at the differences of both Jacobi integrals not only show larger high-frequency errors caused mainly by the kinetic energy but also periodic deviations especially in the potential energy constituents. The reasons for these differences are not clear yet; further investigations are necessary.

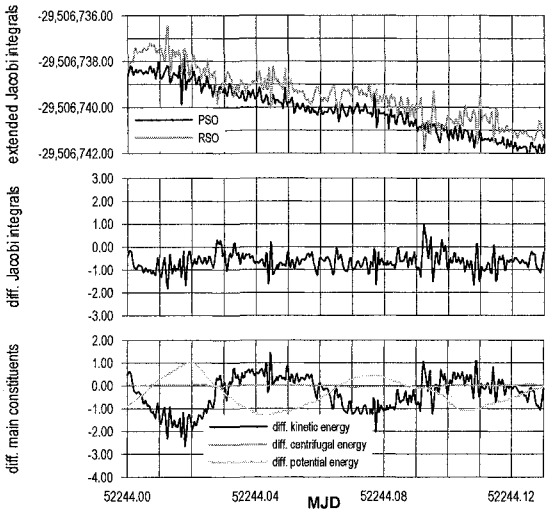


Fig. 5: Comparison of the extended Jacobi integrals (m / s) of a two-revolution PSO and RSO arc of CHAMP (1.12.2001).

6 Conclusions

This investigation demonstrates that the extended Jacobi integral is a valuable procedure to validate the consistency of orbit and force function, which rules the motion of the satellite. Nevertheless, there are some cases where the Jacobi integral is trivially fulfilled: This is the case of dynamic orbit determinations and the case where the force function, mainly the gravity field model, is derived from kinematically determined orbits by applying the energy balance method itself. If the gravity field recovery procedure is not based on the energy integral then it represents a (more or less) independent validation tool. The same holds if the Jacobi integral is based on a gravity field model and satellites orbits, which are not used for the determination of the gravity field model. A critical point is the determination of the velocities from kinematically determined positions. These errors are reflected in the kinetic energy and seem to be larger then the effects of position errors which influence the potential energy. Further investigations are necessary to come up with a careful analysis of these error sources. An important problem to be solved is the separation of space dependent gravity field effects and time dependent orbit effects. The transformation of the deviations from the Jacobi constant into the spectral domain may help to identify various constituents of the deficiencies in the models. The application of the presented validation method to the relative motion of two satellites as in case of GRACE will be treated in a forthcoming paper.

Acknowledgements:

We are grateful to the GeoForschungsZentrum (GFZ) Potsdam for providing the data for this investigation. The support by BMBF and DFG within the frame of the Geotechnologien Programm is gratefully acknowledged.

References:

Bjerhammar, A. (1967): On the energy integral for satellites, Rep. of the R. Inst. of Techn. Sweden, Stockholm, 1967

Falk, G., Ruppel, W. (1973): Mechanik, Relativität, Gravitation, Springer-Verlag, Berlin, Heidelberg, 1973

Förste, Ch., Schwintzer, P., Reigber, Ch. (2001): The CHAMP Data Format, GFZ Internal Publication, CH-GFZ-FD-012

Ilk, K.H. (2002): Energy Relations for the Motion of two Satellites within the Gravity Field of the Earth, in: M.G. Sideris (ed.), Gravity, Geoid and Geodynamics 2000, IAG Symposia Vol. 123, pp. 129-135, Springer-Verlag, Berlin, Heidelberg, 2002

Ilk, K.H. (2003): Energy Relations for Satellite-to-Satellite Motions, in: E.W. Grafarend et al. (ed.), Geodesy – The Challenge of the Third Millenium, pp. 79-91, Springer-Verlag, Berlin, Heidelberg, 2003

Ilk, K.H., Feuchtinger, M., Mayer-Gürr, T. (2003): Gravity Field Recovery and Validation by Analysis of Short Arcs of a Satellite-to-Satellite Tracking Experiment as CHAMP and GRACE, presented at the General Assembly of the IUGG 2003, Sapporo, Japan, 2003

Jekeli, Ch. (1999): The determination of gravitational potential differences from satellite-to-satellite tracking, Celestial Mechanics and Dynamical Astronomy, 7582, pp. 85-100, 1999

Lühr, H., Grunwaldt, L., Förste, Ch., Schwintzer, P., Reigber, Ch., (2001): CHAMP Reference Systems, Transformations and Standards, Internal publication, CH-GFZ-RS-002

McCarthy, D.D. (ed.), (1996): IERS Conventions 1996, Central Bureau of IERS – Observatoire de Paris, Paris, 1996

Montenbruck, O., Gill, E. (2000): Satellite Orbits, Models, Methods and Applications, Springer-Verlag, Berlin, Heidelberg, New York, 2000

O'Keefe, J.A. (1957): An application of Jacobi's integral to the motion of an Earth satellite. The Astronomical Journal, 62(1252), 265-266

Reigber, Ch. (1969): Zur Bestimmung des Gravitationsfeldes der Erde aus Satellitenbeobachtungen, DGK, Reihe C, Heft Nr. 137, München, 1969

Reigber, Ch., (2002): CHAMP Newsletter No. 9, GFZ Internal Publication, Newsletter_CHAMP_009

Reigber, Ch., Balmino, G., Schwintzer, P., Biancale, R., Bode, A., Lemoine, J.-M., Koenig, R., Loyer, S., Neumayer, H., Marty, J.-C., Barthelmes, F., Perosanz, F., Zhu, S.Y., (2002): A high quality global gravity field model from CHAMP GPS tracking data and Accelerometry (EIGEN-1S), Geophysical Research Letters, 29(14), 10.1029/ 2002GL015064

Reigber, Ch., Schwintzer, P., Neumayer, K.-H., Barthelmes, F., König, R., Förste, Ch., Balmino, G., Biancale, R., Lemoine, J.-M., Loyer, S., Bruinsma, S., Perosanz, F., Fayard, T. (2003): The CHAMP-only Earth Gravity Field Model Eigen-2, Advances in Space Research 31(8), 1883-1888, 2003

Springer, T.A., Beutler, G., Rothacher, M., (1998): A new Solar Radiation Pressure Model for the GPS satellites, IGS 1998 Analysis Center Workshop, ESOC, Darmstadt

Biharmonic Spline Wavelets versus Generalized Multi-quadrics for Continuous Surface Representations

Rainer Mautz[1,2], Burkhard Schaffrin[1], Julia Kaschenz[3]

[1] Dept. of Civil and Environ. Eng. and Geodetic Science, Ohio State University, Columbus, Ohio, USA

[2] Institute of Astronomical, Physical and Mathematical Geodesy, Technical University of Berlin, Germany

[3] GeoForschungsZentrum Potsdam, Dept.1: Geodesy and Remote Sensing, Telegrafenberg, Potsdam, Germany

Abstract. In some respect, the continuous representation of surfaces such as the geoid, the topography, or other spatial phenomena, is superior to discrete forms like the TIN (Triangular Irregular Network) or a raster DEM (Digital Elevation Model) as long as these surfaces exhibit a certain degree of local smoothness. In this contribution, we shall concentrate on the special study of biharmonic spline wavelets and of generalized multi-quadrics, with the emphasis on increased efficiency while maintaining the local approximation quality up to the desired resolution.

To support these interpolation techniques a newly developed global search algorithm will be adapted to this problem. It is based on heuristic methods that would allow the user to handle gridded as well as arbitrarily scattered data. The number of coefficients for any surface representation using the global optimization will be extremely small while maintaining a magnitude of deviations that is still acceptable. On the other hand, long computation times have to be taken into account. Concerning the functional model one may face fewer restrictions for its structure when using global methods because no derivatives have to be computed and no approximate values need to be provided.

Some geodetic examples will show the potential of the new techniques, particularly in view of the more classical Fourier analysis.

Keywords. Biharmonic spline wavelets, multi-quadrics, continuous surface representation, global search algorithm

1 Introduction

In geodesy one of the principal research foci is the efficient representation of surfaces. Out of the high number of state-of-the art surface representations only a few can be included in this paper. For a comparison of different methods on a large scale, see Franke (1982). Franke (1987) presents a broad bibliography of 1107 citations related to multivariate approximations. A more recent survey has been carried out by Hubeli and Gross (2000).

All authors mention the difficulty of categorization and the complexity of criteria while trying to rank different methods. Indeed, criteria and their weights depend on the author's emphasis and may include accuracy, timing, storage requirements, ease of implementation, sensitivity to parameters, visual aspects, smoothness / roughness of the surface, ease to compute surface-surface intersections, transitions to different resolutions, possibility of error modeling, behavior at singularities, peaks or cracks, existence of derivatives and gradients. The most important categorization principles differentiate between discrete / continuous, linear / nonlinear and locally / globally supported basis function models.

In the present investigation we shall stick to only four different methods that include the classical Fourier analysis, the multi-quadrics which are well known for their excellent performance, a new family of radial basis functions with local support, and the biharmonic spline wavelets which had been developed for geodetic purposes by Schaffrin et al. (2003). The focus will lie on higher-degree linear versus lower-degree nonlinear models in order to find out how either refinement of the model affects the capability to represent a complex surface.

2 Mathematical Models

"Surface Representations" require the development of continuous surface models from point-sampled data; in effect, it means interpolating and, to some extent, extrapolating point information to cover the continuum. Some mathematical model functions for continuous surface representations are presented.

2.1. a) The 2D-Fourier series:

$$f(x,y)=\sum_{\mu=0}^{\infty}\sum_{\nu=0}^{\infty}A_{\mu\nu}\cos(\mu x)\cos(\nu y) +B_{\mu\nu}\sin(\mu x)\cos(\nu y)+C_{\mu\nu}\cos(\mu x)\sin(\nu y) +D_{\mu\nu}\sin(\mu x)\sin(\nu y) \quad (1)$$

with the frequencies μ, $\nu \in \mathbb{N}_0$ and the amplitudes $A_{\mu\nu}$, $B_{\mu\nu}$, $C_{\mu\nu}$, $D_{\mu\nu} \in \mathbb{R}$. f is a bivariate function from L^2 where the coordinates x and y belong to the interval $[-\pi; +\pi]$. When approximating data $z = f(x, y) + e$, the infinite series is replaced by a summation up to certain integers K_μ and K_ν.

2.1. b) A nonlinear analog to the 2D-Fourier series with flexible frequencies:

$$f(x,y)=\sum_{k=0}^{\infty}A_{\mu_k\nu_k}\cos(\mu_k x)\cos(\nu_k y) +B_{\mu_k\nu_k}\sin(\mu_k x)\cos(\nu_k y)+C_{\mu_k\nu_k}\cos(\mu_k x)\sin(\nu_k y) +D_{\mu_k\nu_k}\sin(\mu_k x)\sin(\nu_k y) \quad (2)$$

The model (2) is no longer harmonic and has been discussed in Kaschenz (2002; 2003) in more detail. It differs from (1) as the frequencies have an extended domain μ_k, $\nu_k \in \mathbb{R}^+$, allowing non-integer ratios between frequencies. However, μ_k and ν_k are unknown parameters, thus making the model (2) *nonlinear*. The series is truncated where the model reaches a predefined precision.

2.2. a) 2-dimensional B-spline wavelets:

$$f(x,y)=\sum_{k_x=1}^{2^{j_{\min}+g}}\sum_{k_y=1}^{2^{j_{\min}+g}}A_{j_{\min},k_x,k_y}\tilde{\phi}_{j_{\min},k_x,k_y}(x,y)+ \sum_{j=j_{\min}}^{J}\sum_{\eta=1}^{3}\sum_{k_x=1}^{2^{j+g}}\sum_{k_y=1}^{2^{j+g}}B_{j,\eta,k_x,k_y}\tilde{\psi}_{j,\eta,k_x,k_y}(x,y) \quad (3)$$

A and B denote the unknown coefficients for the scaling function and the wavelet functions. $\phi(x, y)$ is the 2-D scaling function, $\tilde{\phi}(x, y)$ its dual; $\psi(x, y)$ are the 2-D wavelet functions, and $\tilde{\psi}(x, y)$ their duals. For the reconstruction of the signal only the dual scaling and wavelet functions have to be used, denoted by a tilde. Their polynomial degree is expressed by g. The different levels of detail are denoted by the index j. Wavelet coefficients with a larger j indicate higher detail levels, essentially representing the high-frequency part. The index η denotes the three directional components (horizontal, vertical and diagonal), and the indices k_x, $k_y \in \mathbb{N}_0$ denote the shift of the wavelets to different locations on the (x, y)-patch.

The B-splines on a bounded interval were introduced by Chui and Quak (1992) whereas the related decomposition and reconstruction algorithms were presented in Quak and Weyrich (1994). Via the tensor product at various levels, the 2-D basis functions $\phi(x, y)$ and $\psi_\eta(x, y)$ ($\eta = 1,2,3$) can be generated extremely efficiently. The problem of estimating the coefficients A and B is linear; due to orthogonal subspaces, it is not necessary to solve one big system with linear equations of problem-size, but a sequence of smaller systems. For further details, see Schaffrin et al. (2003).

2.2. b) A nonlinear analog to the 2D B-spline wavelets with flexible dilatation and translation parameters:

$$f(x,y)=\sum_{k=1}^{\infty}B_{j_k,\eta_{lk},x_k,y_k}\psi_{j_k,\eta_{lk},x_k,y_k}(x,y), \quad (4)$$

where $l_k \in \{0, 1, 2, 3\}$. The scaling function ϕ is renamed ψ_0 ($l = 0$) for convenience. To increase flexibility, the dilatation factors $j_k \in \mathbb{R}^+$ are not limited to fixed scales fulfilling 2^j with $j \in \mathbb{N}$. Also the translations x_k and y_k can take any real value in the spatial domain. The directional component of the "pseudowavelets" is kept flexible with the three major directions, allowing $l_k = \{0, 1, 2, 3\}$.

2.3. a) Multiquadrics as radial basis functions with global support:

Out of the large number of radial functions, we pick the well known multiquadrics, introduced by Hardy (1971). Numerical results can be found in Hardy and Hou (1990). Multiquadrics have been a subject of comparison in Franke (1982). Hales and Levesley (2000) introduced a hierarchical approach to determine the coefficients efficiently. Multiquadrics are defined as the upper sheet of a hyperboloid of revolution. Their functional model reads

$$f(x,y)=\sum_{k=1}^{n}A_k\sqrt{r_k^2+c^2} \quad (5)$$

where r_k is given by

$$r_k^2(x,y)=(x-x_k)^2+(y-y_k)^2 \quad (6)$$

as the radial distance between the evaluation point (x, y) and a fixed center position (x_k, y_k) which could be the location of one of the observations. A_k are the unknown parameters and $c \in \mathbb{R}$ is a predefined constant. The estimation of the unknown parameters A_k is a linear problem. An alternative to the multiquadric function is its reciprocal,

$$f(x,y)=\sum_{k=1}^{n}A_k\left(r_k^2+c^2\right)^{-\frac{1}{2}} \quad (7)$$

also known as inverse multiquadric. Figures 1-3 show the radial multiquadric-function for various choices of c, and Figure 4 the inverse multiquadric-function.

2.3. b) Radial basis functions, "multiquadrics" with flexible positions:

$$f(x,y)=\sum_{k=1}^{n} A_k\sqrt{r_k^2+c_k^2}. \tag{8}$$

The model (8) is similar to (5) except that $x_k \in \mathbb{R}$ and $y_k \in \mathbb{R}$ are unknowns. Also the parameter c is here introduced as an unknown c_k for every $k \in \{1, 2, ..., n\}$. Thus the model is more flexible, but nonlinear and has to be treated as such.

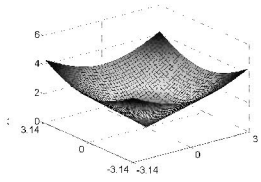


Fig. 1 Multiquadric function with $c = 0$, $A = 1$ and the center in the origin. The surface is a cone.

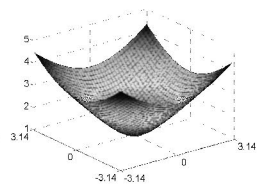


Fig. 2 Multiquadric function with $c = 1$, $A = 1$ and the center in the origin.

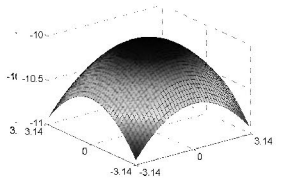


Fig. 3 Negative multiquadric function with $c = 10$, $A = -1$ and the center in the origin.

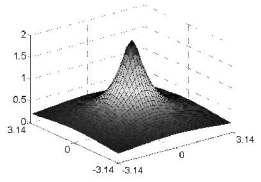


Fig. 4 Inverse multiquadric function with $c = 1$, $A = 0.5$ and the center in the origin.

2.4. a) Radial basis functions with local support, linear model:

When the basis functions differ from zero only within a certain distance from the center point, the represented function depends mainly on the local data. The cut-off can be performed in such a way that there is no discontinuity in the function and in its first derivative. The model (9) with fixed center positions (x_k, y_k) and parameters c_k was considered.

2.4. b) Radial basis functions with local support, nonlinear model:

Again, the nonlinear model

$$f(x,y)=\begin{cases}\sum_{k=1}^{n} A_k\left(\frac{r^2}{c_k^2}-1\right)^2, & r<c_k\\ 0, & r\geq c_k.\end{cases} \tag{9}$$

is introduced by considering the center positions (x_k, y_k) and the parameters c_k as unknowns for every $k \in \{1, 2,..., n\}$.

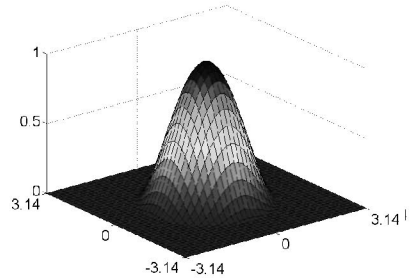


Fig. 5 Radial basis function with local support; Model (9): Polynomial, truncated at $r = c$ with $c = 2.0$, $A = 1$.

3 Solution Techniques for the Linear and Nonlinear Models

Let us work with the linear model

$$\mathbf{y}=\mathbf{A}\xi+\mathbf{e},\ \mathbf{e}\sim\left(0,\sigma_0^2\mathbf{P}^{-1}\right),\ \mathbf{P}:=\mathbf{I}_n,\ \operatorname{rank}(\mathbf{A})=m<n, \tag{10}$$

where $\mathbf{y}$ is a $n\times 1$ vector of observations (sample of function values), $\mathbf{A}$ is a $n\times m$ matrix of known coefficients, ξ is a $m\times 1$ vector of unknown parameters (typically the "amplitudes" of the basis functions), and $\mathbf{e}$ is the $n\times 1$ random error vector. All observations are assumed to be uncorrelated and of same weight, thus as weight matrix $\mathbf{P}$, the $n\times n$ identity matrix $\mathbf{I}_n$ is chosen. Using BLUUE (Best Linear Uniformly Unbiased Estimator), we obtain the estimated parameters $\hat{\xi}$, the residuals $\tilde{\mathbf{e}}$, and the estimated variance component $\hat{\sigma}_0^2$ with respective variances and covariances:

$$\hat{\xi}=\mathbf{N}^{-1}\mathbf{c}\sim\left(\xi,\sigma_0^2\mathbf{N}^{-1}\right),\quad [\mathbf{N},\mathbf{c}]:=\mathbf{A}^{\mathrm{T}}\mathbf{P}[\mathbf{A},\mathbf{y}],$$

$$\tilde{\mathbf{e}}=\mathbf{y}-\mathbf{A}\hat{\xi}\sim\left(0,\sigma_0^2\left(\mathbf{P}^{-1}-\mathbf{A}\mathbf{N}^{-1}\mathbf{A}^{\mathrm{T}}\right)\right), \tag{11}$$

$$\hat{\sigma}_0^2=(n-m)^{-1}\left(\mathbf{y}^{\mathrm{T}}\mathbf{P}\mathbf{y}-\mathbf{c}^{\mathrm{T}}\hat{\xi}\right)=\tfrac{\tilde{\mathbf{e}}^{\mathrm{T}}\mathbf{P}\tilde{\mathbf{e}}}{n-m}.\ \hat{\mathbf{D}}\left\{\hat{\sigma}_0^2\right\}=\tfrac{2\hat{\sigma}_0^2}{n-m}.$$

The systems for solving the unknowns in capital letters (e.g., the amplitudes A, B, C, D) are linear if there are no further parameters. In this case, all the position vectors, scaling factors and frequencies are assumed to be known. In fact, they all are set to be constant at certain values, usually equidistant measures that allow efficient algorithms or multi-scaling representations. In the literature, not much effort has been documented to introduce the positions, scaling coefficients and frequencies as unknowns. The resulting models become nonlinear, and the solving techniques require the high computational effort of global optimization methods. Local, nonlinear adjustment strategies like the Gauss-Newton iteration depend on the provision of sufficiently reliable starting values. If they cannot be provided as required, the problem is of global type when applied to the least-squares target function. The common

global techniques such as heuristic methods, interval strategies or genetic algorithms have been sufficiently discussed elsewhere. For instance, global models with unknown frequencies in time series have been studied by Mautz (2001; 2002) while global models for surface data were discussed by Kaschenz (2002; 2003).

4 Comparison of the Models

Due to the high number of potential criteria, a universal ranking of different model functions is impossible. Instead of attempting to weigh different criteria to establish one specific ranking, only the major criteria of the different models will be compared in order to show some of their advantages and drawbacks.

4.0. Premises

Some models necessitate a special data structure for efficient handling. The Fourier series (1) and wavelet analysis (3) use equidistant data points. In addition, multiresolution representations like the wavelet model require a special grid size (e.g. 2^j by 2^j, $j \in \mathbb{N}$). Thus, scattered observations have to be adjusted onto a grid in order to apply these models. The resulting approximation errors from gridding have not been taken into account in this study.

The following comparisons are based on the specific dataset shown in Figure 6.

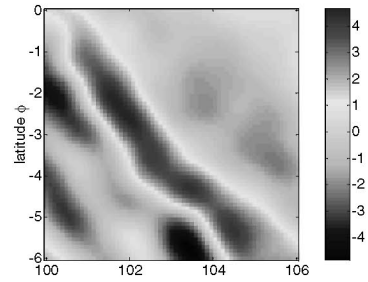


Fig. 6 Height anomalies in [m] from EGM 96, degree and order > 70, between 100° and 106° longitude and 0° to -6° latitude, and sampled onto a 65×65 grid. The range of the data values is [–4.8 m; +4.7 m].

4.1. Adaptation of the model

The residual information is essential for many criteria, which can be expressed by numbers quite well. However, it should not be neglected that the out-

Table 1: Comparison of various surface representation models with a constant number of unknowns; units: [m] or [m^2]

Model	Predetermined coefficients	No. of basis functions	Number of unknowns	C_1: $\max\{\lvert\tilde{e}_j\rvert\}$	C_2: $\sum_{j=1}^{n}\lvert\tilde{e}_j\rvert/n$	C_3: $\tilde{\mathbf{e}}^T\mathbf{P}\tilde{\mathbf{e}}$	C_4: $m_0=\sqrt{\hat{\sigma}_0^2}$
2D-Fourier (linear)	μ, ν	81	289	2.28	0.13	340	0.29
2D-Fourier (nonlinear)	μ_k, $\nu_k \in$ [0; max]	48	288	0.61	0.05	20	0.07
Wavelet, 4 levels	scale of j_0	289	289	0.42	0.05	25	0.08
Wavelet, compression	scale of j_0	289	289 (+ 578)	0.32	0.04	13	0.06
multiquadric (linear)	$c = 0.5°$	289	289	0.34	0.03	8	0.04
multiquadric, flex. pos.	$c = 6°$	96	288	0.80	0.08	46	0.11
multiquadric, c & pos.flex.	$c \in [0°; 1°]$	72	288	0.65	0.08	49	0.11
inv. multiquadric., (linear)	$c = 0.5°$	289	289	0.37	0.04	15	0.06
local radial function	$c = 2°$	289	289	0.68	0.09	70	0.13
loc. rad. func., pos. & c flex.	$c \in [0.05°; 5°]$	72	288	0.23	0.05	14	0.06

Table 2: Number of parameters necessary to achieve a given degree of approximation (in terms of rms)

Model		$m_0=\sqrt{\hat{\sigma}_0^2}$ [m]	$\max\{\lvert\tilde{e}_j\rvert\}$ [m]	No. of basis functions	No. of unknowns	Relative redundancy [%]
2D-Fourier	linear	0.08 m could not be reached. The minimal m_0 = 0.28m (625 unknowns).				
2D-Fourier	nonlinear	0.072	0.61	48	288	93.18
Tensor Wavelet, 4 levels	linear	0.080	0.42	289	289	93.16
Tensor Wavelet, compressed	linear	0.080	0.44	220	220 + (440 integers)	94.80 (89.59)
Multiquadrics (c =0.5°)	linear	0.080	0.51	225	225	94.67
Multiquadrics	nonlinear	0.079	0.48	118	354	91.62
inv. Multiquadrics (c =0.5°)	linear	0.077	0.56	225	225	94.67
local radial function (c = 2°)	linear	0.075	0.46	484	484	88.54
local radial function	nonlinear	0.080	0.29	56	224	94.70

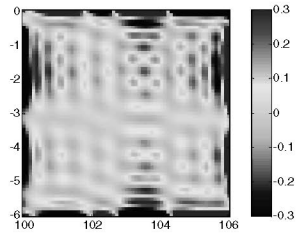


Fig. 7 Residuals of the 2D-Fourier series model (1); 289 basis functions. The unit is [m].

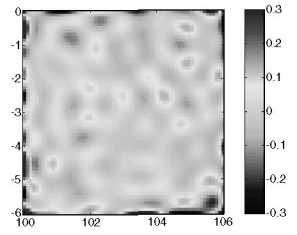


Fig. 8 Residuals of the nonlinear analog to the Fourier series (2); 48 basis functions. The unit is [m].

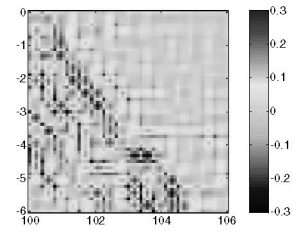


Fig. 9 Residuals of the linear B-Spline Wavelet model (3); 289 basis functions including levels 0 - 4. The Unit is [m].

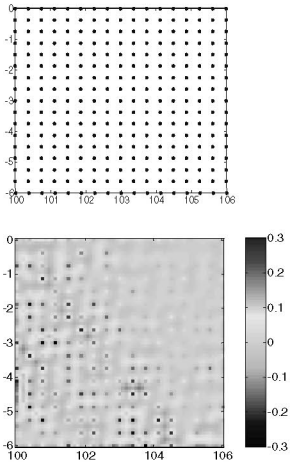


Fig. 10 a-b Linear multiquadric model (5); 289 basis functions with $c = 0.5$. Location of the center points and residuals. The unit is [m].

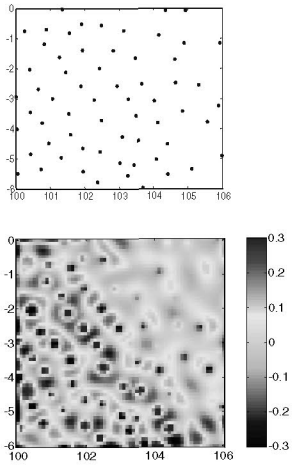


Fig. 11 a-b Nonlinear multiquadric model (8); 72 basis functions with flexible positions and variable parameter c. Location of the center points and residuals. The unit is [m].

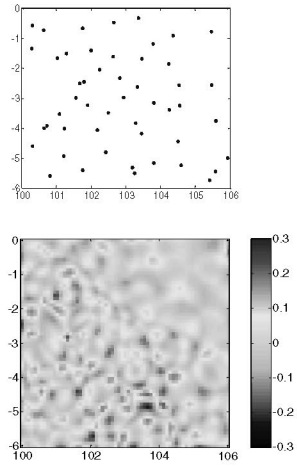


Fig. 12 a-b Nonlinear radial functions with local support (9); 72 basis functions with flexible positions and variable parameter c. Location of the center points and residuals. The unit is [m].

come of such an evaluation strongly depends on the properties of the data. In order to rate the residual information, the following criteria were used:

$$C_1 := \max\left\{\left|\tilde{e}_j\right|, j = 1,2,...,n\right\}, \quad C_2 := \sum_{j=1}^{n}\left|\tilde{e}_j\right| / n,$$

$$C_3 := \tilde{\mathbf{e}}^{\mathrm{T}}\mathbf{P}\tilde{\mathbf{e}}, \quad C_4 := \sqrt{\hat{\sigma}_0^2} = \sqrt{\tilde{\mathbf{e}}^{\mathrm{T}}\mathbf{P}\tilde{\mathbf{e}}/(n-m)}.$$

4.2. Timing

The computational effort depends mainly on the number of observations n and the number of parameters m. In general, linear models involve the solution of a system of $O(m)$ equations. Therefore, it is necessary to set up a system of normal equations $O(n\ m^2)$ and perform an inversion $O(m^3)$. However, there are some exceptions: Orthogonal Fourier basis functions on the continuum - model (1) - would create a diagonal normal matrix, thus the complexity reduces to $O(n\ m)$. The B-spline wavelets - model (3) - are semi-orthogonal and have local support, causing banded matrix structures. Furthermore, the size of the equation systems can be reduced by a factor $m^{-1/2}$. Making use of these properties, the computational effort is $O(n\ m^{1/2})$. The multiquadrics - model (5) - are of complexity $O(n\ m^2)$. If a hierarchical data thinning algorithm is used according to Hales and Levesley (2000), the algorithm may be of linear complexity, but only if the data are spaced equidistantly. However, when the radial basis functions have local support - model (9) - the computational burden decreases drastically. These functions are suitable to handle very large m and n, when globally supported radial functions may not be feasible.

The computing time for nonlinear models is longer than the time for linear models since a global algorithm may require several times to solve a system of linear equations. The computational effort for a nonlinear system will mainly depend on (a) the number of unknown parameters, (b) the possibility to set up a linear system if a fraction of the parameters are kept constant, (c) the global strategy and its implementation, (d) the degree of adaptation

to the problem on hand, (e) the number of local minima of the objective function, (f) the desired precision, and (g) whether the global solution has to be guaranteed. A more detailed discussion can be found in Xu (2002; 2003).

4.3. Smoothness of the surface

The behavior of six different models, namely (1), (2), (3), (5), (8), (9), is visualized in Figures 7-12, for a patch between 100° to 106° longitude and 0° to –6° latitude respectively. All models have 288 (or 289) parameters. As shown in Figure 7 and Table 1, the Fourier series does not represent the data well. In contrast, the nonlinear analog in Figure 8 shows very small residuals in the center.

The wavelet model shows distinct horizontal features. At the center point locations of the multiquadric models, some peaks can be seen.

The locally supported radial basis functions show a smooth surface throughout.

5. Conclusions

In general, linear and local models are computationally efficient. The shortest computation time was obtained for the linear wavelet model.

Using the 2D-Fourier series – model (1) –, one cannot reach the goal of 8 cm rms deviation because, with an increase in the number of parameters, the model improves only very slightly. With the decrease of redundancy the Fourier model reaches the minimal rms of 0.28 cm when $K_\mu = K_\nu = 13$ and the number of parameters is 625. The nonlinear Fourier model has an unexpectedly poor maximum deviation despite a reasonable mean deviation. The linear 2D-Fourier and wavelet models need a data-grid for efficient handling.

The parameter c in the multiquadric model is not very sensitive to the data. Thus c = const. is reasonable. Linear radial basis models need a pre-defined value for c. In order to obtain satisfying results, c has to be optimized prior to their application.

The multiquadric models show peaks, and the wavelet models horizontal and vertical edges in their representations as well as in their residuals.

Nonlinear models generally need a smaller number of basis functions, particularly the locally supported radial functions. Although very time consuming, the smallest number of unknowns was used here to achieve a rms deviation of 8 cm as well as the smallest maximum deviation, but with an enlarged mean deviation when compared with the linear multiquadrics. Note that the nonlinear wavelet model (5) still needs to be studied.

Acknowledgements

R. Mautz acknowledges the partial support of a Feodor-Lynen scholarship of the Alexander-von-Humboldt Foundation (Germany).

References

Chui, C.K., E. Quak (1992): *Wavelets on a bounded interval*; in: Numerical Methods in Approximation Theory, Vol. 9, (D. Braess and L. Schumaker, eds.), Birkhäuser: Basel, Boston, Berlin, pp. 53-77.

Franke, R. (1982): *Scattered data interpolation: Tests of some methods*; Maths. Comp. 38, pp. 181-199.

Franke, R. (1987): *Recent advances in the approximation of surfaces from scattered data*; in: Multivariante Approximation (C.K. Chui, L. Schumaker, and F. Utreras, eds.), Academic Press, New York, pp. 79-98.

Hales S., J. Levesley (2000): *Multi-level approximation to scattered data using inverse multiquadrics*; in: Curve and Surface Fitting: Saint-Malo 1999, (A. Cohen, C. Rabut and L. Schumaker, eds.) Vanderbilt Univ. Press, Nashville/TN, pp. 247-254.

Hardy, R. (1971): *Multiquadric equations of topography and other irregular surfaces*; J. Geophys. Res. 76, pp. 1905-1915.

Hardy, R., Z. Hou (1990): *A new comparative test and a report on the new multiquadric biharmonic method*; AVN-International Edition 7/1990, pp. 4-17.

Hubeli, A., M. Gross (2000): *A survey of surface representations for geometric modeling*; CS Technical Report #335, February 28, 2000, Inst. of Scientific Computing, Swiss Federal Institute of Technology, Zurich, Switzerland.

Kaschenz, J. (2002): *Multidimensional spectral analysis with problem-oriented frequencies* (in German); Dept. of Geodesy and Geoinformation, Tech. Univ. of Berlin, Diploma Thesis.

Kaschenz, J. (2003): *Data modelling using various frequency series* (in German); Z. für Vermessungswesen (ZfV) 128, pp. 260-265.

Mautz, R. (2001): *On the determination of frequencies in time series solving nonlinear adjustment problems;* German Geodetic Comm., C-532, Munich, Germany.

Mautz, R. (2002): *Solving nonlinear adjustment problems by global optimization*; Bollettino di Geodesia e Scienze Affini, Vol. 61, No.2, pp. 123-134, 2002.

Quak, E., N. Weyrich (1994): *Decomposition and reconstruction algorithms for spline wavelets on a bounded interval*; Applied and Comput. Harm. Analysis 1, pp. 217-231.

Schaffrin, B., R. Mautz, C.K. Shum, H. Tseng (2003): *Towards a spherical pseudo-wavelet basis for geodetic applications;* Computer-Aided Civil and Infrastructure Engineering 18, pp. 369-378.

Xu, P. (2002): *A hybrid global optimization method: the one-dimensional case*; Journal of Comp. and Applied Mathematics 147, pp. 301-314.

Xu, P. (2003): *A hybrid global optimization method: The multi-dimensional case*; Journal of Comp. and Applied Mathematics 155, pp. 423-446.

An Energy Approach for Intelligent Structural Health Monitoring Using Wavelet Analysis

N. El-Sheimy
Department of Geomatics Engineering,
The University of Calgary, 2500 University Dr. N.W., Calgary, Alberta T2N 1N4, Canada

M. M. Reda Taha
Department of Civil Engineering, University of New Mexico, Albuquerque, NM, USA 87131-0001

A. Osman
Department of Electrical Engineering
The University of Calgary, 2500 University Dr. N.W., Calgary, Alberta T2N 1N4, Canada

A. Noureldin
Department of Electrical and Computer Engineering,
Royal Military College of Canada, P.O. Box 17000 Station Forces, Kingston, Ontario, Canada

Abstract: The premise in using principles of vibration-based damage identification in structural health monitoring (SHM) is that when damage occurs, the mechanical properties of the structure will be altered and thus the vibration response of the structure to externally applied loads and environmental conditions will be changed. The low frequency of civil structures and the limited effect of local damage on vibration characteristics of these structures make the development of efficient SHM systems challengeable.

This paper introduces the Wavelet multi-resolution analysis as a promising technique for damage identification. Analytical analysis of simulated sensor signals significantly corrupted by other noise signals showed that the wavelet multi-resolution analysis was able to separate the different frequency components and to identify the original simulated structural response. Combining wavelet multi-resolution analysis with artificial neural networks makes it possible to pattern structural dynamics during healthy performance and detect changes of the structural dynamics when damage is introduced to the system.

Keywords: Wavelet transform, Structural health monitoring, Damage detection.

1 Introduction

Recent surveys of the civil infrastructure assets in Canada and the US have indicated that about 40% of North American infrastructure urgently requires replacement or rehabilitation. Structural health monitoring (SHM) systems can accurately monitor the response of structures due to real-time loading and environmental conditions. In spite of the high initial cost, SHM systems are expected to be able to enhance structural safety and result in considerable future cost savings through early detection of the need for maintenance. Thus, the purpose of utilizing SHM systems is to detect damage occurrence in the structure and potentially inform the responsible authorities of the location and nature of this damage upon its occurrence (Kinawi et al. 2002). Modal analysis derived from analyzing the output signal of a set of sensors distributed in the structure (e.g. Modal Assurance Change (MAC) implemented by means of Fourier transform is presently the damage detection technique mostly utilized (Gentile 2003). Experimental verification of the algorithms on real structures showed that modal characteristics (i.e. Fourier transform) might be insensitive to localized damage (Doebling et al. 1998). It has also been proven that changes in the mode shapes due to damage might not be identifiable due to signal corruption by noise and/or due to the limited number of identifiable mode shapes (Doebling et al. 1998).

Therefore, the use of wavelet multi-resolution analysis (WMRA) is examined here as a potential means of separating the different frequency components in the output signals of several inertial-based sensors distributed over the structure. WMRA offers a time-resolution representation of the signals. This means that the different frequency components of a signal can be analyzed within a localized area of the whole signal. Therefore, wavelet analysis can de-noise the sensor signal without considerable degradation and can detect equally the rapidly changing high-frequency

(details) components as well as the slowly changing low-frequency (approximations) components of the signal. The decomposition process can be iterated, with successive approximations being decomposed in turn so that the inertial sensor signal is broken down into many low-resolution components that permit accurate pattern recognition.

2 Digital Signal Processing of Structural Dynamics

2.1. Fourier Transform

Fourier transform (FT) has been traditionally used for signal processing of stationary signals. This method can be used to decompose a function in terms of a set of basis functions in the form of a set of complex sinusoids (Strang, G. and Nguyen). Because FT is the result of integration (or sum in the discrete domain) over the whole signal length, the resultant coefficients cannot indicate the time at which a specific frequency existed in the signal. FT thus provides perfect frequency information, but no time information (Ogden 1997).

In order to reduce the effect of non-stationary signals on FT, short-time FT (STFT) assumes local periodicity within a continuously translated time window (Strang, G. and Nguyen). The difficulty in analyzing a non-stationary signal lies in determining the moment in time at which a given transient occurred. Time localization is essential since the frequencies in a non-stationary signal do not exist for all time throughout the signals. If it were possible to divide the non-stationary signals into a number of smaller, stationary signals then the FT of each sub-division is accomplished by applying a window function to the input signal before computing the FT. As the width of the window function decreases, a smaller portion of the input signal is considered, guaranteeing greater time localization of the frequency components in the signal (Ogden 1997). As the width of the window function increases, more accurate information about frequencies within the window is obtained. However, the ability to determine the time within the input signal at which those transients occurred is lost (Ogden 1997). In the discrete domain, the implementation of STFT uses a square wave window that truncates the sine or cosine function to fit a window of a particular width. Because a single window is used for all frequencies in the STFT, the resolution of the analysis is the same at all locations in the time-frequency plane.

Although the STFT provides a time-frequency representation of a signal, there is a major drawback with respect to utilizing STFT in SHM applications: this is because the window width is fixed. Uncertainty principles indicate that it is not possible to determine both the frequency components and at time at which these components exist within a signal (Daubechies 1988). The time-frequency tradeoff corresponds to the choice of the window function's width. A narrow window width leads to good time localization but poor frequency localization. On the other hand, a wide window width leads to poor time localization but good frequency resolution. Therefore, fine time resolution for short duration and high frequency signals, and fine frequency resolution for long duration and lower frequency signals are non-attainable using STFT (Robertson et al. 1996).

2.2 Wavelet Transform (WT)

The continuous wavelet transform (CWT) of a continuous signal x (t) is defined as:

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\, \psi\left(\frac{t-b}{a}\right) dt \qquad (1)$$

where a and b are known as the dilation and translation parameters and $\psi(t)$ is called the analyzing function (Strang and Nguyen 1997). $\psi(t)$ is not limited to the complex exponential function as for FT. In fact, the only restriction on $\psi(t)$ is that it must be short and oscillatory; it must have zero average and decay quickly at both ends (Daubechies 1988). This restriction ensures that the integral in equation (1) is finite and gives the name wavelet or "small wave" to the transform. $\psi(t)$ is referred to as the "mother wavelet" and it dilates and translates simply as "wavelets" or "daughters wavelets". The dilation parameter "a" represents the scale, which determines the oscillatory behaviour of the particular daughter wavelet, and the translation parameter "b" represents the shifting of the mother wavelet, which is important for the transform to provide time-localization information in the original signal (Strang and Nguyen 1997).

In the discrete-time domain, the implementation of the wavelet transform is based on a bank of discrete time filters Thus WMRA decomposes signals into various resolution levels (Strang and Nguyen 1997). It is suggested herein that WMRA will be applied to the individual signals received from the different inertial-based sensors (e.g. accelerometers, inclinometers and possibly MEMS)

distributed over the structure. The filtering approach of WMRA is to form a series of half-band filters that divide a spectrum into high frequency and low frequency bands. It is formulated on a scaling function or LP filter and wavelet function or HP filter. The WMRA builds a pyramidal structure (as shown in Fig. 1) that requires an iterative application of scaling and wavelet functions to low pass and high pass filters, respectively (Strang and Nguyen 1997). The high frequency band outputs are taken as the detail coefficients (D1, D2, D3), and the low frequency band outputs are taken as the approximation coefficients (A1, A2, A3).

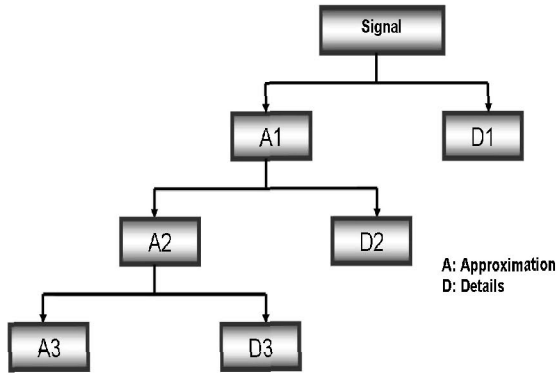


Fig. 1 The pyramid structure of WMRA

3 Application to Structural Health Monitoring Systems

The responsibility of the diagnostic module in the structural health monitoring (SHM) systems is to detect damage occurrence (Doebling et al. 1998). In intelligent SHM systems, the diagnostic module has to provide feature extraction and pattern recognition of the system dynamics (Kinawi et al. 2002). The ability of WMRA to separate the major frequency components of the sensed signals in the time domain gives it the potential to be an efficient tool for damage detection in SHM systems. The combination of the WMRA and artificial intelligence paradigms make it possible to extract features from the structural dynamics and recognizing pronounced changes in these features subsequently for damage identification. This is achieved by using neural networks to pattern the structural dynamics during healthy performances and to predict then at later instances of unknown performance to detect damage occurrence.

In this context, the WMRA is iterated, with successive approximations being decomposed in turn, so that the signal is broken into many fine resolution components. The original signal is then reconstructed from the sum of the final approximation and detail components of all levels after excluding those signals that represent noise. Therefore WMRA is utilized at two levels: firstly to filter the sensor output so that de-noised "clean" signals representing structural dynamics can be attained; and secondly to permit comparison between the acquired and predicted signals at various levels of decomposition.

3.1 Wavelet Neural Network module

The bridge response due to the dynamic load was recorded and used as to train the Wavelet-Neural Network module named (WA_NN) designed to de-noise, pattern and predict the system dynamics (Reda Taha et al. 2003). A schematic representation

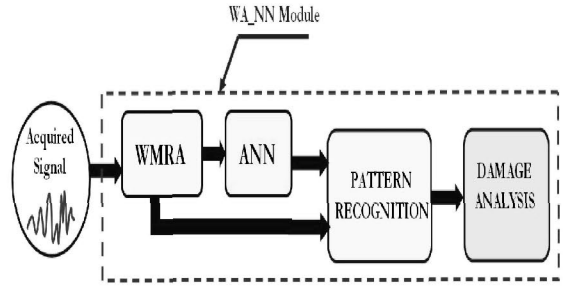


Fig. 2 Schematic representation of the WA_NN module.

Noise was introduced to the entire simulated signal to represent the effect of the sensing and communication systems on the signal at the data acquisition level. The measurements were noised using a random signal before the simulated measurements were forwarded to the WA_NN module. Afterwards, the bridge dynamic responses (accelerations) were de-noised using WMRA by applying principles of soft thresholding (Reda Taha et al. 2003).

The re-constructed signal was then forwarded to the WA_NN module to recognize patterns in the system dynamics. The WA_NN is a multi-layer feed forward artificial neural network developed to predict the system dynamics. While linear normalization algorithms were used for normalizing the system input parameters, linear and log-sigmoid functions were used as transfer functions connecting the hidden and external layers. The network was trained to achieve a mean square error (MSE) of $8*10^{-4}$. The network was then used to predict the dynamic response in "healthy"

conditions. The predicted and "acquired" signals were then decomposed using WMRA to facilitate signal comparisons at the decomposed levels.

3.2 Wavelet norm index for damage occurrence detection

Consider the signal f(t) at different resolution levels and present it as a series expansion by using a combination of scaling function $\varphi(t)$ and wavelet function $\psi(t)$ as presented in equation (2):

$$f(t)=\sum_{k} c\ (k)\varphi_k(t)+\sum_{k}\sum_{j=0}^{J-1} d(j,k)\psi_{j,k}(t) \tag{2}$$

where J represent the total number of resolution levels and the expansion coefficients c(k) represent the approximation of the original signal f(t) with a resolution of one point per every 2J-1 points of the original signal and the expansion coefficients d(j, k) represent details of the original signal at different resolution levels. The coefficients c(k) and d(j, k) are calculated by inner product (Gouda et al. 2002) as:

$$c(k) = <\varphi_k | f> = \int_{-\infty}^{\infty} f(t)\varphi_k(t)\,dt \tag{3}$$

$$d(j,k) = <\psi_{j,k} | f> = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)\,dt \tag{4}$$

In the wavelet domain, the wavelet coefficients C_{mc} represent the acquired signal and C_{pc} represent the predicted signal. The coefficients of error signal between the predicted and the acquired signals in the wavelet domain can be obtained by subtracting the wavelet coefficients of the predicted signal from the corresponding wavelet coefficients of the acquired signal at the different resolution levels as in equation (5):

$$C_{ec} = C_{mc} - C_{pc} \tag{5}$$

If the approximation and detail coefficients used form an orthonormal basis, Parseval's theorem relates the energy of the error signal to the norm of the wavelet coefficients (Mallat 1999). Thus, the norm of the error signal at any instance "N" can be expressed in terms of the wavelet coefficients of the predicted and acquired signals. The wavelet norm index of the error signal (I_{eN}) is equal to the square of the norm of the error wavelet expansion coefficients and can be mathematically represented as:

$$I_{eN} = \left[\sum_{k} \left| C_{(J-1)ec}(k) \right|^2 + \sum_{k}\sum_{j=1}^{J-1} \left| d_{(j)ec}(k) \right|^2 \right] \tag{6}$$

The differentiation between the damaged and healthy conditions can thus be noticed by evaluating the wavelet norm index of the error signal (I_{eN}) at various operating instances. As the neural networks are trained to predict the system dynamics under healthy conditions, I_{eN} shall be relatively constant during healthy operations. When damage is induced in the structure, significant changes in the value of I_{eN} is expected. The expected change in the I_{eN} value is attributed to the fact that damage results in the presence of new superimposed dynamic components in all subsequent response upon damage occurrence.

4 Probability of Damage Occurrence

By observing the statistical characteristics of the wavelet norm index of the error signal (I_{eN}), the probability of damage occurrence can be estimated. The mean ($\overline{I_{eM}}$) and the variance ($\overline{\sigma_{eM}}^2$) of the wavelet norm index of the error signal can be established after observing a set of M healthy population measurements at priori-known healthy instances (after training period). The probability of damage occurrence at an unknown instance (K) is the probability that the wavelet norm index of the error signal at instance K (I_{eK})is greater than the mean wavelet norm index $\overline{I_{eM}}$.

As the wavelet norm index of the error signal can be assumed normally distributed, the probability of damage occurrence can be evaluated using the cumulative Gaussian normal distribution function as in equations (7) and (8):

$$z = \frac{\left| E_{eK} - \overline{E_{eM}} \right|}{\sigma_{EM}} \tag{7}$$

$$P_D(K) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2}\,dz \tag{8}$$

where $P_D(K)$ is the probability of damage occurrence at the instance K. By establishing a

threshold value for the probability of damage occurrence, the NN_WA module will identify damage occurrence if the mean value for the probability of damage for a set of consecutive instances exceeded the damage threshold value.

5 Case Study: Detecting Simulated Damage of a Concrete Bridge

5.1 The model

A three-dimensional finite element (FE) model of a 40m span simply supported, prestressed concrete bridge was developed. The bridge consists of 4 girders spaced at 4.0 m. Each girder has a trapezoidal cross section with a 1.65 m depth and a mass per unit length of 4800 kg/m. The characteristic concrete strength was 45 MPa with a modulus of elasticity of 34700 * 10^3 MPa. The bridge model consisted of 60 beam elements and 40 shell elements. The model includes twenty one (21) nodes for each girder with three degrees of freedom (DOF). The degrees of freedom (DOF) of the FE nodes included the translations along the X and Z axis and rotation about the Y axis. End bearings were modelled as supports. The analysis assumed linear elastic behaviour of the bridge. A schematic of the bridge is presented in Fig. 3. A set of 30 accelerometers was assumed to be distributed over the bridge to monitor accelerations due to traffic loading. Simulated output signals of these accelerometers are obtained from the FE model as the z-axis acceleration component at the nodes located at the accelerometer locations.

The features of the structural dynamics of the bridge during "healthy" performance were extracted by analyzing the dynamic response of the bridge under different traffic loads. Truckloads were modelled as passing over the bridge at different speeds ranging between 20 km/hr and 70 km/hr. The dynamic effect of the traffic loads was assumed to occur when the truck passed over an irregularity in the asphalt deck. Deck irregularities were randomly distributed along the bridge span. The dynamic effect followed a trapezoidal function with a time period of 15 seconds and a peak occurring for 5 seconds.

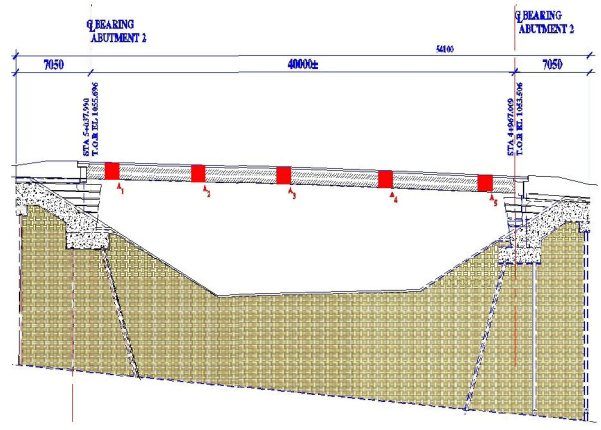


Fig. 3 The 40 m span, prestressed single span concrete bridge

5.2 Evaluating the probability of damage occurrence

The dynamic responses of the bridge in two states of "health" were examined: undamaged or "healthy" and damaged. Damage was introduced to the FE model by reducing the cross sectional stiffness of three elements to 65% of the original value. Each signal was monitored in real-time and was compared to its counterpart signal predicted by the WA_NN module. Comparison between acquired and predicted signals was performed at the different levels of decomposition suing Daubechies wavelet function (Daubechies 1998). Sample comparison between simulated and acquired signal at a damage instance is shown in Fig. 4.

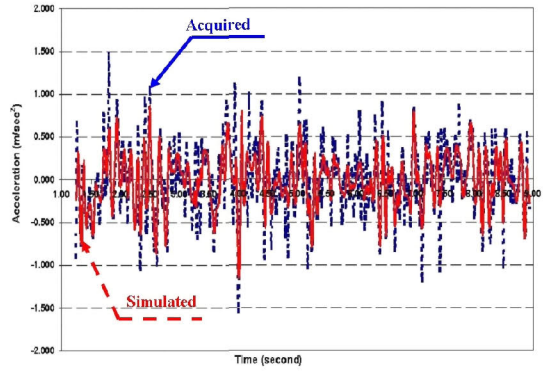
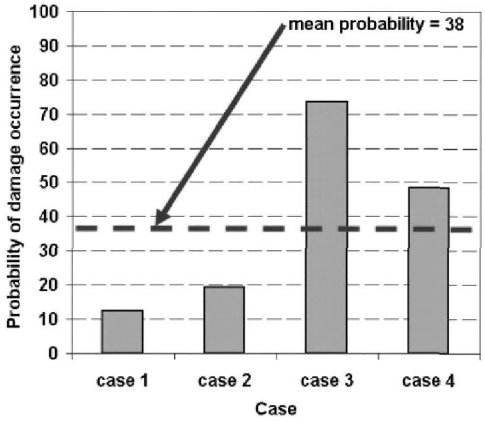


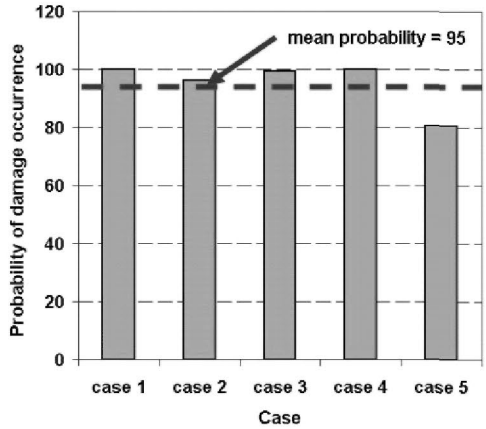
Fig. 4 Comparison between acquired and simulated signals at damage instance.

The WA_NN module was able to detect the structural health condition of both instances very accurately. The WA_NN module predicted a mean probability of damage occurrence during the instance of healthy operational of 38% versus a

mean probability of damage occurrence during the instance of unhealthy operational of 95%. The second case obviously exceeded the damage threshold and thus and an alert will be sent to the off-site server indicating a highly probable damage occurrence in the bridge. Fig. 5 [a] and [b] show the difference in the probability in damage occurrence at both the healthy and damage instances.



[a] Healthy condition



[b] Damaged condition

Fig. 5 Acquired and predicted reconstructed signals in healthy and damaged conditions.

5 Conclusions

The combination of wavelet multi-resolution analysis (WMRA) and artificial intelligence makes it possible to extract features from real time signals and to establish a model that can pattern and recognize these features.

The proposed system was able to detect damage occurrence successfully from signals simulated using a finite element model of a prestressed concrete bridge.

6 Acknowledgement

The financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC) is gratefully acknowledged.

References

Daubechies, I. (1988) Orthogonal bases of compactly supported wavelets. *Communications in Pure Applied Math*, 41(7), pp. 909-996.

Doebling, S. W., Farrar, C. R., and Prime, M. B. (1998). A Summary review of vibration-based damage identification methods. *The Shock and Vibration Digest*, 30 (2), pp. 91-105.

Gentile, C. (2003). Dynamic performance of A P.C. bridge: computation and tests. *Proceedings of the ICPCM Conference,* Cairo, Egypt, El-Dieb et al. Eds., Vol. 1, pp. 449-458.

Gaouda A. M. A.M. Kanoun, S.H.; Salama, M.M.A.; Chikhani, A.Y. (2002). Wavelet-based signal processing for disturbance classification and measurement. *IEEE proceedings: Generation, transmission and distribution*, 149 (3) pp. 310-318

Kinawi, H., Reda Taha, M. M. and El-Sheimy, N. (2002). Structural health monitoring using the semantic wireless web: A novel application for wireless networking. *Proceedings of the 27th IEEE Conference (LCN),* FL, USA, pp. 770-780.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transaction on Pattern analysis and Machine intelligence*, 11 (7): 674-693.

Ogden, R.T. (1997). Essential Wavelets for Statistical Applications and Data Analysis, Birkhäuser Boston, USA.

Reda Taha, M. M, Noureldin, A. Osman, A. and El-Sheimy, N. (2003). An introduction to the use of wavelet multi resolution analysis for intelligent structural health monitoring. *In Review, Canadian Journal of Civil Engineering*.

Robertson, D.C. Camps, O.I. Mayer J.S. and Gish, W.B. (1996). Wavelets and electromagnetic power system transients. *IEEE Transaction on Power Delivery*, 11(2), pp. 1050-1056.

Strang, G. and Nguyen, T. (1997) Wavelet and Filter Banks, Wellesley-Cambridge Press, NY, USA.

On Discrete Schemes in Downward Continuation of Gravity

Wenke SUN
Earthquake Research Institute, University of Tokyo, Tokyo, Japan
sunw@eri.u-tokyo.ac.jp

Abstract. Two discretization schemes of the Poisson integral have been proposed to date. Although they are mathematically solvable, they produce different gravity on the geoid for the same input data. This discrepancy arises because of different discretization techniques of the Poisson kernel; still, this problem has not received adequate attention. The question is whether the system of the discretization is reasonable. Methods to discretize the Poisson integral are investigated in this study. For this purpose, a single mean scheme is presented to evaluate numerically the Poisson integral. A comparison between the point and mean schemes shows that, for a limit topographical grid size, the point discretization scheme results in a serious theoretical problem since it greatly underestimates gravity on the geoid, and even gives incorrect results for extreme cases. A careful construction of the coefficient matrix for the discrete system is much more important than using point gravity as input.

Keywords. Gravity, Downward continuation, Poisson integral, Discretization scheme, Geoid

1. Introduction

For geoid determination by any boundary-value problem of potential theory the disturbing gravity potential (and all its functions) must be harmonic everywhere above the geoid. For this reason observed values of gravity are corrected for gravitational effects of all masses above the geoid. From existing approaches for gravity reduction that harmonize observed gravity, the Helmert second method of condensation is often used. It is recommended to use such a method that not only harmonizes but also smoothes the gravity field (Heiskanen and Moritz, 1967; Wichiencharoen, 1982; Vaníček and Martinec, 1994; Martinec and Vaníček, 1994). The downward continuation can be applied to the disturbing potential and its functions or even their linear combinations.

Once a physical model is selected, the remaining work is left to mathematics. It is known that the downward continuation of gravity cannot be performed for the continuous gravity field on the real irregular topography (Heiskanen and Moritz, 1967). The downward continuation for discrete functions can generally be performed without difficulty (Bjerhammar, 1962). This implies that one has to discretize the Poisson integral to a system of linear equations so that the gravity data on the geoid can be obtained by solving this system. Bjerhammar (1962 and 1987) studied this problem using a point scheme. This point scheme was also adopted by, e.g., Martinec (1996).

Vaníček et al. (1996) and Sun and Vaníček (1998) studied the downward continuation of 5'x 5' mean Helmert's gravity anomalies. They discretized the Poisson integral using a double mean scheme. They claimed that the evaluation of the downward continuation of the mean 5'x 5' Helmert gravity anomalies is a well-posed problem with a unique solution; it can be carried out routinely to any accuracy desired in geoid determination. They proved that a discretized system of Poisson integral equations can always be solved for the 5'x 5' discretization. However, numerical investigations indicated that the double mean scheme involved tedious numerical work.

On the other hand, although the above two systems of linear equations are mathematically tractable, they produce different results for the same input (Martinec, 1996; Sun and Vaníček, 1998). The main reason for the difference between the two discretization schemes is assumed to arise from the discretization of the Poisson kernel. Detail discussions on the treatment of the numerical discretization can be found in Huang (2002).

This paper investigates possible discretization methods of the Poisson integral. A single mean scheme is proposed to overcome the disadvantage of the double mean scheme. A comparison between the point, single and double mean schemes shows that the mean schemes are efficient and reasonable, but the point scheme presents a serious theoretical problem. Although the point scheme is numerically solvable, gravity obtained on the geoid is greatly

underestimated. Actually, the difficulty of dealing with the Poisson integral is due to the behavior of the Poisson kernel K. This research shows that a careful construction of a coefficient matrix of the discrete system of the Poisson integral is much more important than using point gravity as an input.

2. Poisson's integral and existing discretization schemes

We start with the Poisson integral (spherical approximation)

$$g_t = \int_{\Omega} K(H,\psi) g_g d\Omega \qquad (1)$$

where g_t is gravity observed on the Earth's surface, g_g is gravity on the geoid, $d\Omega$ stands for the surface element, H is the topographical height, ψ is the angular distance between the computation and moving points, and $K(H,\psi)$ is the Poisson kernel function given by

$$K(H,\psi) = \frac{R^2}{4\pi r}\frac{r^2 - R^2}{l^3} \qquad (2)$$

where R is the mean radius of the earth, r is the radial distance of the calculating point, and $r=R+H$, l is the spatial distance between the calculating and moving points.

Theoretically, the Poisson integration in Eq. (1) should be performed over the entire spherical surface of radius R. However, since the Poisson kernel $K(H,\psi)$ attenuates rather quickly with the growing angular distance ψ (see Fig. 1), integration can be truncated at a relatively small angular distance ψ_0. Also, the truncation errors can be either estimated or neglected.

Since gravity g_t is observed on topography and gravity g_g is required on the geoid, solving Eq. (1) is an inverse problem. If g_t is known continuously everywhere on the irregular topography, the downward continuation of gravity in Eq. (1) is still an ill-posed problem (Heiskanen and Moritz, 1967). However, if gravity is observed or given in a discrete form, i.e., Eq. (1) is solved discretely, it can always be determined uniquely. In practice, since one cannot observe gravity continuously on the Earth's surface, one must solve Eq. (1) in a discrete manner.

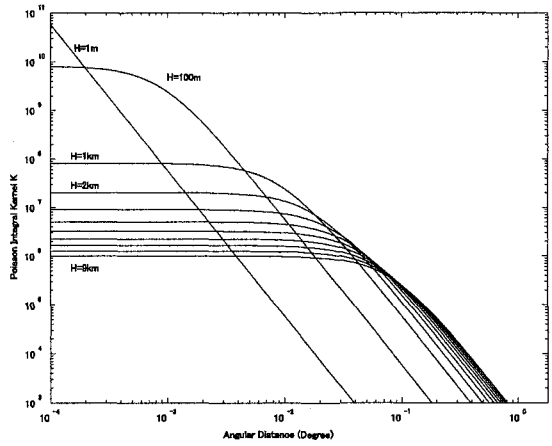


Fig. 1. Poisson's integral kernel for heights from 1 m to 9 km.

For this purpose, the point scheme can be used (e.g., Bjerhammar, 1962 and Martinec, 1996), so that the Poisson integral is simply discretized to the system of linear equations:

$$g_{ti} = \sum_j K_{ij} g_{gj} \Delta S_j \qquad (3)$$

or in vector form

$$\mathbf{g_t} = \mathbf{B}\mathbf{g_g} \qquad (4)$$

where the elements of the coefficient matrix **B** are given by

$$B_{ij} = K(H,\psi_{ij})\Delta S_j \qquad (5)$$

where ψ_{ij} is the angular distance between centers of the i-th and j-th cells, and ΔS_j is the area of the i-th cell.

For an area with high topography, (say, H>3 km), the Poisson kernel is relatively small and the off-diagonal elements of the matrix **B** are significant. This implies that observed gravity over a large area is strongly involved. In this case, gravity on the geoid can be obtained by the inverse of the discrete system in Eq. (4), i.e.,

$$\mathbf{g}_g = \mathbf{B}^{-1}\mathbf{g}_t \qquad (6)$$

When H is small, say, H<1 m, the Poisson kernel

approaches a δ-function (refer to Fig. 1). In this case, the coefficient matrix **B** can be simply replaced by its diagonal elements, since all off-diagonal elements are very small and can be neglected; thus, Eq. (6) becomes

$$\mathbf{g}_g = \{B_{ii}\}^{-1}\mathbf{g}_t \tag{7}$$

Vaníček et al. (1996) and Sun and Vaníček (1998) proposed the double mean scheme to deal with the discretization problem. The idea was based on the following consideration: The final product of the downward continuation is to be used in producing the geoid by Stokes's convolution integration, which is also evaluated by replacing integration by a summation over mean anomalies on a grid of geographical cells. It is thus natural to use the same geographical grid for both tasks, i.e., to compute the downward continuation of gravity with the same mean anomalies as those used in the Stokes convolution. Therefore, taking the average of gravity anomalies, the discrete Poisson integral for the mean anomaly in the i-th cell (c_i) in the grid becomes

$$\overline{g_{ti}} = \frac{1}{\Delta S_i}\sum_j \int_{c_i}\int_{c_j} \overline{g_g} K(H,\psi)d\Omega' d\Omega \tag{8}$$

where the $\overline{g_t}$ and $\overline{g_g}$ are mean gravity on the topography and geoid, respectively, ΔS_i is the area of the i-th cell. Since mean gravity $\overline{g_g}$ is constant in the j-th cell (c_j), it can be moved outside the integral, so that, Eq. (8) becomes

$$\overline{g_{ti}} = \sum_j \overline{g_{gj}}\,\overline{\overline{B_{ij}}} \tag{9}$$

$$\overline{\overline{B_{ij}}} = \frac{1}{\Delta S_i}\int_{c_i}\int_{c_j} K(H,\psi)d\Omega' d\Omega \tag{10}$$

In vector form, Eq. (9) can be rewritten as

$$\overline{\mathbf{g}_t} = \overline{\overline{\mathbf{B}}}\,\overline{\mathbf{g}_g} \tag{11}$$

This treatment is called the double mean discretization scheme because of the double integral of the Poisson kernel in Eq. (10).

The term "Mean scheme" has a two-fold meaning: the mean value of gravity and the Poisson kernel. The mean gravity value is considered to be better than the point value because it removes a high-frequency gravity signal and smoothes the gravity field over the geographical grid. Actually, even though point gravity is used, once the difference of $Max\left|g_t^{point} - g_t^{mean}\right|$ is small enough, the used mean or point gravity value is equivalent. However, compared to the mean gravity value, the mean Poisson kernel is more important. It directly concerns how to correctly discretize the Poisson kernel, or how to correctly construct coefficient matrix **B**. Discussing the problem of discretization is the main task of this research.

3. Single mean scheme

For convenience, we consider another way to discretize the Poisson integral – the single mean scheme. Since observed gravity g_t in each geographical cell is given in practice by its mean value (for point gravity, we can still assume it as the mean value), i.e., $\overline{g_t}$ is constant, then gravity g_g on the geoid can be assumed to be the mean value – constant $\overline{g_g}$ inside the cell. When the Poisson integral in Eq. (1) is discretized, the mean gravity anomaly $\overline{g_{gj}}$ can be moved outside the integral: i.e.,

$$\overline{g_{ti}} = \sum_j \overline{g_{gj}} \int_{c_j} K(H,\psi)d\Omega \tag{12}$$

where c_j is the area of the j-th geographical cell. Equation (12) can be expressed in vector form as

$$\overline{\mathbf{g}_t} = \overline{\mathbf{B}}\,\overline{\mathbf{g}_g} \tag{13}$$

where coefficient matrix $\overline{\mathbf{B}}$ contains a single integral of the Poisson kernel over the cell c_j:

$$\overline{B_{ij}} = \int_{c_j} K(H,\psi)d\Omega \tag{14}$$

So far, we have defined three discretization schemes to deal with the Poisson integral: the point, simple mean and double mean schemes. Mathematically, these systems are quite simple

containing only two parts: observed gravity and the coefficient matrix. As mentioned above, point gravity and mean gravity can be assumed to be the same: i.e., $\overline{\mathbf{g}}_t = \mathbf{g}_t$, $\overline{\mathbf{g}}_g = \mathbf{g}_g$. This assumption is reasonable in practice because in many cases, it is impossible to observe gravity at many points on the Earth's surface; even gravity for a non-observed area is estimated by interpolation. In this case, a point value is usually used for the mean value.

A possible explanation for the different results in the downward continuation is the difference between the coefficient matrices $\mathbf{B}$, $\overline{\mathbf{B}}$ and $\overline{\overline{\mathbf{B}}}$. Assuming that the systems of linear equations in Eqs. (4), (11) and (13) are well-posed and the high-frequency portions of gravity anomalies are smoothed, one can solve these systems and obtain convergent numerical results. However, this does not mean that the results are correct. The problem is whether or not the discretization of the Poisson integral is correct, or the system is valid. This problem has not received much attention. It is known that taking the average of gravity smoothes the high frequency portion; at the same time, the mean value is considered to represent well the gravity field in a geographical cell. For the same reason, the discretized Poisson kernel should well represent its real behavior in the same cell because the kernel in the vicinity of the computation point varies more than gravity anomalies (see Fig. 1). In the following sections the three discretization schemes are compared while considering the validity of the coefficient matrices.

Compared to the point scheme, the two mean schemes can be basically considered identical. Numerical investigations show that the double mean scheme does not improve the coefficient by much over the single one. However, from an application point of view, the single mean scheme is much more convenient than the double one, since numerical calculation of the coefficient matrix of the latter is very tedious. Since both single and double mean schemes require at least one integration of the Poisson kernel, and significantly differ from the point scheme, the following discussion mainly addresses the point and single mean schemes.

4. Principle of discretization

A basic principle for the valid discretization of any integral is that the discrete kernel function value should accurately represent the true (integrated) value in each grid or cell. In other words, the difference between discrete and true values should be sufficiently small; otherwise discretization will bring a serious error to the result. In the following discussion, this principle is used to evaluate the accuracy of a discretization.

We first consider a general case, assuming an integral of an arbitrary one-dimensional function $f(x)$ multiplied by a step function $g(x)$ that is constant over each division Δx_i within the integration limits x_1 and x_2, i.e.,

$$I(f) = \int_{x_1}^{x_2} g(x) f(x) dx \qquad (15)$$

Discretizing the integral in Eq. (15) using the point scheme leads to a quadrate summation over divisions Δx_i, which can be written as

$$\tilde{I}(f) = \sum_i g(x_i) f(x_i) \Delta x_i \qquad (16)$$

The function $g(x_i)$ does not cause any problem in the discretization process since it is constant over Δx_i. However, the function $f(x_i)$ plays a key role in Eq. (16), depending on its behavior. For example, if $f(x)$ is a δ-function for an extreme case. Of course, the discretization scheme in Eq. (16) cannot be applied for the δ-function; however, the example at least indicates that the behavior of the function $f(x)$ should be carefully considered, because when Δx_i is constant, the function $f(x)$ is very large at the central cell (refer Fig. 1), i.e.,

$$f(x) >> 1 \qquad (17)$$

and then Eqs. (15) and (16) give different results:

$$I(f) = \int_{x_1}^{x_2} g(x) f(x) dx = g(x_0) \qquad (18)$$

$$\tilde{I}(f) = \sum_i g(x_i) f(x_i) \Delta x_i >> g(x_0) \qquad (19)$$

The above discussion indicates that discretization schemes of the integral in Eq. (15) depend not only on $g(x)$, but also on $f(x)$. In case of the Poisson integral, we may replace $g(x)$ and $f(x)$ by g_g and

$K(H,\psi)$, respectively. Notice that the Poisson kernel $K(H,\psi)$ is much more complicated than sin x, as shown in Fig. 1, since integration of the Poisson kernel yields (Sun and Vaníček, 1998)

$$\int_{\psi<1^o} K(H,\psi)d\Omega \approx \int_{\Omega} K(H,\psi)d\Omega = 1 \quad (20)$$

A good approximation of Eq. (20) is, of course, to guarantee that the coefficient matrix is about one; that is, for the point scheme

$$\sum_{j=1}^{N} K(H,\psi_{ij})\Delta S_j \approx 1 \quad (21)$$

where N is the total number of geographical cells within a spherical cap of the 1° radius. We can examine the converging behavior of the sum when N is increasing. in Eq. (21). Results for different topographical heights are calculated and plotted in Fig. 2. It can be seen that for n<100

$$\sum_{j=1}^{N} K(H,\psi_{ij})\Delta S_j >> 1 \quad (22)$$

This indicates that a rather large number N is required if the relation in Eq. (21) should hold true. In other words, we can achieve an accurate estimation of the Poisson kernel using a point discretizing scheme only when the geographical cell is sufficiently small. For example, for the topographical height of 4.5 km, N should be taken to be at least 100. For the height of 100 m, N=1000 is still insufficient. Consequently, this behavior of the simple sum in Eq. (21) implies that if the point discretization scheme is adopted, a huge number of linear equations must be solved, which is usually very difficult or even impossible.

However, one problem is that the cell size is usually predetermined according to the density of observed gravity data, such as 5'x 5'. Hence, the point scheme does not guarantee accuracy of Eq. (21). In this case, the mean scheme presents an advantage since, for any cell size, the following relation holds true

$$\int_{\Omega} K(H,\psi)d\Omega = \sum_{j=1}^{N} \int_{c_j} K(H,\psi)d\Omega = 1 \quad (23)$$

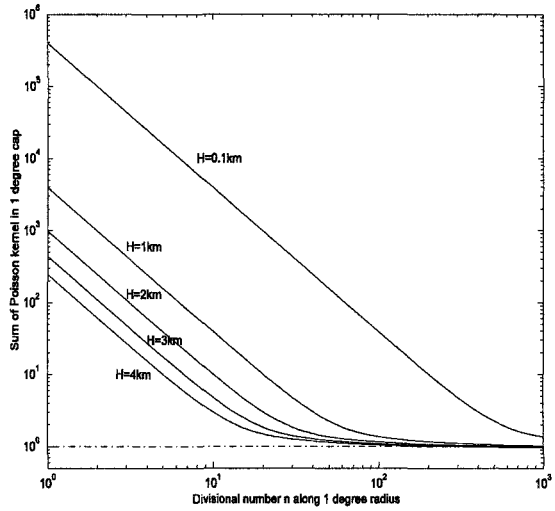


Fig. 2. Converging behavior of the sum of the discretized Poisson kernel as the divisional number n increases, for different H.

If the sum in Eq. (21) is carried out over the 1° spherical cap, we can still obtain a reasonable approximation, i.e.,

$$\sum_{j=1}^{N} \int_{c_j,\psi<1^o} K(H,\psi)d\Omega \approx 1 \quad (24)$$

The truncation error in Eq. (24) depends on the topographical height H. Results show that the truncation error is relatively small.

Unlike the sine function, the Poisson kernel behaves like the δ-function, especially for small topographical heights (Fig. 1). This means that the diagonal term of the matrix $\mathbf{B}$ or $\overline{\mathbf{B}}$ always dominates with respect to the off-diagonal terms. Hence, it is simple to investigate mean and point values for a central cell since it concerns the diagonal terms. For this purpose, discretization is assumed again to be carried out for a fixed 5'x 5' grid scheme. Then, the diagonal terms for the mean and point schemes are written as

$$\overline{B_{ii}} = \int_{c_i} K(H,\psi)d\Omega \quad (25)$$

$$B_{ii} = K(H,\psi_i)\Delta S_i \quad (26)$$

Numerical results of $\overline{B_{ii}}$ and B_{ii} for a few heights are calculated and listed in Table 1; it is apparent that the values of B_{ii} are much larger than those of

$\overline{B_{ii}}$, i.e.,

$$B_{ii} >> \overline{B_{ii}} \qquad (27)$$

The differences between $\overline{B_{ii}}$ and B_{ii} can be described by their ratio ($B_{ii}/\overline{B_{ii}}$), see Table 1. The ratio can reach rather large values, which indicates that the point scheme will greatly under estimates the gravity on the geoid. Therefore, it may be concluded that the point scheme for the Poisson integral does not satisfy the discretization principle; it will bring a considerable error to the downward continuation of gravity.

Table 1. Comparison of the diagonal terms of the coefficient matrices $\mathbf{B}$ and $\overline{\mathbf{B}}$.

H(m)	$\overline{B_{ii}}$ (H)	B_{ii} (H)	($B_{ii}/\overline{B_{ii}}$)
1	1.00	4.93e6	4.93e6
100	0.89	492.0	555.6
1000	0.47	4.92	10.5
2000	0.29	1.23	4.24
3000	0.19	0.546	2.88
4000	0.14	0.308	2.20

5. Conclusions

So far, two discretization schemes of the Poisson integral have been presented and unique numerical results have been obtained. However, for different schemes, different gravities are obtained on the geoid even though the inputs are the same. The reason for this discrepancy is different discretization techniques for the Poisson kernel. This problem has not received sufficient attention. This research presents a new single mean scheme to implement the Poisson integral. Numerically, the single and mean schemes give almost the similar numerical results, but the single one is numerically simpler and considerably reduces numerical calculation. The comparison of the point and mean schemes shows that the point scheme significantly underestimates gravity; it also gives an incorrect result for the extreme case. In conclusion, the point scheme can introduce a large error to the downward continuation of gravity, but the mean schemes are reasonable and recommended.

Acknowledgments. Useful comments on the first draft of the manuscript by Dr P. Xu are highly appreciated. Comments by Dr. J. Janak and an anonymous reviewer are also gratefully acknowledged. This research was financially supported by JSPS research grant "Basic design and feasibility studies for future missions for monitoring Earth's environment".

References

Bjerhammar, A (1962) On an explicit solution of the gravimetric boundary value problem for an ellipsoid surface of reference. Stockholm, Royal Institute of Technology, Geodesy Division

Bjerhammar A (1987) Discrete Physical Geodesy, Report No. 380, Department of Geodetic Science and Surveying. The Ohio State University, Columbus, Ohio

Heiskanen WH and Moritz H (1967) Physical Geodesy. Freeman and Co., San Francisco

Huang, J (2002) Computational methods for the discrete downward continuation of the earth gravity and effects of lateral topographical mass density variation on gravity and the geoid, PhD thesis, University of New Brunswick.

Martinec Z and Vaníček P (1994) Direct topographical effect of Helmert's condensation for a spherical approximation of the geoid. Manuscr Geod 19: 257-268

Martinec Z (1996) Stability investigations of a discrete downward continuation problem for geoid determination. J of Geodesy 70: 805-828

Moritz H (1980) Advanced Physical Geodesy, Wichman, Karlsruhe

Sun W and Vaníček P (1998) On some problems of the downward continuation of the 5'X5' mean Helmert gravity disturbance. J of Geodesy, 72: 411-420

Vaníček P and Martinec Z (1994) The Stokes-Helmert scheme for the evaluation of a precise geoid. Manuscr Geod 19: 119-128

Vaníček P, Sun W, Ong P, Martinec Z, Vajda P, and Horst B (1996) Downward continuation of Helmert's gravity. J of Geodesy, 71: 21-34

Wichencharoen C (1982) The indirect effects on the computation of geoid undulations. Dept. Geod. Sci. Rep. 336, The Ohio State University, Columbus

Application of Clebsch-Gordan Coefficients and Isomorphic Frame Transformations to Invert Earth's Changing Geometrical Shape for Continental Hydrological Loading and Sea Level's Passive Response

[1,2]Geoffrey Blewitt, [2]Peter Clarke, [1]David Lavallée and [2]Konstantin Nurutdinov
[1]Nevada Bureau of Mines and Geology, and Seismological Laboratory, University of Nevada, Reno, U.S.A.
[2]School of Civil Engineering and Geosciences, University of Newcastle, Newcastle upon Tyne,, U.K.

Abstract. The hydrological cycle induces mass exchange between the oceans and continents, and redistribution of water within the continents. In turn, relative sea level passively responds to the change in shape of both the geoid and the ocean bottom. The time variation in these two surfaces must self-consistently relate to the total mass redistribution. Here we show formally how, if given a time series of station coordinates from a global network of geodetic stations, we can invert for the separate contributions to the total mass redistribution due to water on the continents and water in the oceans. If we characterize the deformation of the solid Earth in terms of a truncated vector spherical harmonic expansion, the solution to this problem can be formulated in terms of Clebsch-Gordan coefficients, which are familiar for the calculation of branching ratios in elementary particle physics. As part of a rigorous solution to this problem, we formally address the definition of reference frame by applying the concept of an "isomorphic frame transformation," which conserves the formal theoretical relationship between surface deformation and load distribution by a simultaneous translation of origin and change of degree-one load Love numbers.

Keywords. Clebsch-Gordan, isomorphic, hydrological cycle, frame transformation, Love number

1 Introduction

We have developed a methodology to invert for the redistribution of fluids on the Earth's surface given precise global geodetic measurements of Earth's geometrical shape [*Blewitt and Clarke*, 2003]. Three steps are involved: (1) inversion of geodetic station coordinates for a spherical harmonic representation of Earth's shape; (2) inversion of Earth's shape for surface mass distribution; (3) inversion for a specific surface mass distribution consistent with static ocean equilibrium theory.

Here we focus on the methodology of step (3), which is facilitated by the use of Clebsch-Gordan (CG) coefficients. The method is applied where it is appropriate to spatially partition a spherical function into two spatial domains (land and ocean) which are characterized by different properties (such as the land's ability to concentrate water).

Secondly, we highlight our findings on reference frame theory, showing how to transform loading equations between frames [*Blewitt*, 2003].

2 Clebsch-Gordan Coefficients

2.1 Spectral Model for Function on Sphere

An arbitrary function on the sphere $f(\Omega)$ at position Ω (latitude φ, longitude λ) can be expanded:

$$f(\Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{nm} Y_{nm}(\Omega) \tag{1}$$

where the complex spherical harmonics are expressed using associated Legendre polynomials:

$$\begin{aligned} Y_{nm}(\Omega) &= \sqrt{\frac{2l+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_{nm}(\sin\phi) e^{im\lambda} \\ &= (-1)^m Y^*_{n(-m)}(\Omega) \end{aligned} \tag{2}$$

where

$$\oiint Y_{nm}(\Omega) Y^*_{n'm'}(\Omega) d\Omega = \delta_{nn'} \delta_{mm'} \tag{3}$$

Therefore coefficients in equation (1) for a real-valued function are given by

$$f_{nm} = \oiint f(\Omega) Y^*_{nm} d\Omega = (-1)^m f^*_{n(-m)}(\Omega) \tag{4}$$

2.2 Spectral Model for Multiple Functions

Consider the case where it is appropriate to express

$$f(\Omega) = \tilde{f}(\Omega)c(\Omega) \tag{5}$$

We shall be interested in the case where $\tilde{f}(\Omega)$ is a function that varies (with choice of model, epoch, or data set) and is represented as a truncated expansion to degree $\bar{n}$, and $c(\Omega)$ is a constant function (that does not change with scenario) known to any arbitrary degree and order. In the language of signal processing,, we might consider $\tilde{f}(\Omega)$ as the signal and $c(\Omega)$ the filter. Applying equation (4):

$$\begin{aligned} f_{nm} &= \oint\!\!\!\!\!\int \tilde{f}(\Omega)c(\Omega)Y^*_{nm}(\Omega)d\Omega \\ f_{nm} &= \sum_{n',m'}^{\bar{n}} \tilde{f}_{n'm'} \left(\sum_{n'',m''}^{2n} c_{n''m''} Q_{n'm';n''m'';nm} \right) \end{aligned} \tag{6}$$

where we have chosen to first sum over the coefficients of the constant function $c(\Omega)$, and where

$$Q_{n'm';n''m'';nm} \equiv \oint\!\!\!\!\!\int Y_{n'm'}(\Omega)Y_{n''m''}(\Omega)Y^*_{nm}(\Omega)d\Omega \tag{7}$$

Particle physics and the theory of atomic spectra enjoy broad application of equation (7), which can be computed in terms of tabulated CG coefficients (in angled brackets) [*Particle Data Group*, 2002]:

$$\begin{aligned} Q_{n'm';n''m'';nm} = &\sqrt{\frac{(2n'+1)(2n''+1)}{4\pi(2n+1)}} \\ &\times \langle n',m';n'',m''|n,m\rangle \langle n',0;n'',0|n,0\rangle \end{aligned} \tag{8}$$

2.3 Selection Rules

The general result of equation (8) is zero unless the following well-known "selection rules" are obeyed:

$$\begin{aligned} m &= m' + m'' \\ |n' - n''| &\le n \le n' + n'' \\ n + n' + n'' &= \text{even} \end{aligned} \tag{9}$$

These rules significantly limit the number of summation terms in equation (6). "Selection rules" are fundamental in quantum phenomena as they specify which possible outcomes of an interaction are allowed or forbidden. Such rules derive from the properties of spherical harmonics. For example, the multiple of two pure degree-2 spherical harmonics is a linear combination of only degrees 0, 2, and 4.

A useful consequence of selection rules for our purposes is that $c(\Omega)$ need only be expressed to degree $2n$ to get an *exact* result in equation (6). This fact can be exploited, for example, if we associate $c(\Omega)$ with the ocean function, which is defined to be one over the oceans and zero on land, and is therefore discontinuous on coastlines:

$$c(\Omega) \equiv \begin{cases} 1 (\Omega \in \text{ocean}) \\ 0 (\Omega \in \text{land}) \end{cases} \tag{10}$$

Ocean function coefficients are computed using coastline data [*Balmino et al.*, 1973]. It may be counter-intuitive that the ocean function can be accounted for exactly in equation (6) even though in principle such a discontinuous function is poorly represented (alone) by a truncated expansion.

2.4 Product-to-Sum Operator

Given that $c(\Omega)$ is independent of scenario, it is convenient in equation (6) to first tabulate (once and for all) the following matrix of coefficients:

$$C_{nm;n'm'} = \sum_{n'',m''}^{2n} c_{n''m''} Q_{n'm';n''m'';nm} \tag{11}$$

for all $n' \le \bar{n}$ and all n of potential interest. Once tabulated, equation (6) becomes the matrix equation

$$\begin{aligned} f_{nm} &= \sum_{n',m'}^{\bar{n}} C_{nm;n'm'} \tilde{f}_{n'm'} \\ \mathbf{f} &= \mathbf{C}\tilde{\mathbf{f}} \end{aligned} \tag{12}$$

where matrix $\mathbf{C}$ is the "product-to-sum operator." Thus the problem of multiplying two functions in the spatial domain has been converted to a matrix multiplication of spectral coefficients. In the case that $c(\Omega)$ is the ocean function, equation (12) computes the filtering effect of the ocean mask on the spectral coefficients. This allows relative sea level to be represented as a lower-degree spherical harmonic function than would otherwise be required to account for the coastlines.

3 Spatial Representation Model

3.1 Naïve Mapping

Consider the general situation for a function $f(\Omega)$ we have determined by some experimental means the spectral coefficients up to degree and order $\bar{n}$, and we then wish to map this function on the sphere. In this case the "naïve" spatial representation of $f(\Omega)$ is a truncated spherical harmonic expansion:

$$f(\Omega) \cong \bar{f}(\Omega) = \sum_{n,m}^{\bar{n}} f_{nm} Y_{nm}(\Omega) \tag{13}$$

3.2 Spatial Domain Partitioning

Suppose we know *a priori* that $f(\Omega)$ is discontinuous at coastlines, and perhaps we have a model that distinguishes its behavior separately over the ocean and on land. It is better to incorporate the *a priori* information by partitioning the spatial representation between the ocean $s(\Omega)$ and land $l(\Omega)$:

$$\begin{aligned} f(\Omega) &= s(\Omega) + l(\Omega) \\ &= c(\Omega)\tilde{s}(\Omega) + [1 - c(\Omega)]\tilde{l}(\Omega) \end{aligned} \tag{14}$$

where $c(\Omega)$ is the ocean function, and where the tildes denote functions that are smooth with no discontinuity at the coastlines.

Now we can apply the product-to-sum conversion given by equation (12) to find the relationship between spectral coefficients of $f(\Omega)$ and the spectral coefficients of the smooth functions over ocean $\tilde{s}(\Omega)$ and land $\tilde{l}(\Omega)$:

$$\begin{aligned} f_{nm} &= \sum_{n',m'} C_{nm;n'm'} \tilde{s}_{n'm'} \\ &\quad + \sum_{n',m'} (\delta_{nn'}\delta_{mm'} - C_{nm;n'm'}) \tilde{l}_{n'm'} \end{aligned} \tag{15}$$

$$\mathbf{f} = \mathbf{C}\tilde{\mathbf{s}} + (\mathbf{I} - \mathbf{C})\tilde{\mathbf{l}}$$

Equation (15) is the spectral equivalent to equation (14). For example, given low-degree approximations to land and ocean loading models, equation (15) can be used to compute the spectrum of the total load accounting for effect of the coastline discontinuity. The spatial equivalent to this computation would be to take equation (14) and apply a Green's function convolution [*Wahr*, 1982].

There are advantages to the spectral approach, however, when investigating long-wavelength phenomena, or for constructing an invertible system of equations to estimate low-degree loading coefficients directly from observational data. For example, equation (15) may be more convenient for studying the effect of degree-2 loading on Earth rotation, or degree-1 loading on geocenter motion.

3.3 Physical Constraints

Consider the case where the smooth functions over ocean and land are functionally related through the physics of the problem (e.g., through the shape of the geoid and the shape of the deformed Earth). If *a priori* physical constraints can be applied such that the number of free variables are equal on both sides of equation (15), it then becomes possible to invert this for ocean and land functions truncated at degree $\bar{n}$. In Section 4 we give concrete examples of how to apply such physical constraints.

3.4 Partitioned Spatial Representation

The last step is to then map these smooth functions only over their respective domains, that is with discontinuities at the coastlines.

$$f(\Omega) \cong \hat{f}(\Omega) = \sum_{n,m}^{\bar{n}} \begin{cases} \tilde{s}_{nm} Y_{nm}(\Omega \in \text{ocean}) \\ \tilde{l}_{nm} Y_{nm}(\Omega \in \text{land}) \end{cases} \tag{16}$$

Equation (15) guarantees that the coefficients of $f(\Omega)$ are consistent with the original truncated expansion, but in effect, the spatial solution (16) augments the truncated expansion to infinite degree and order so as to be consistent with the physical relationship between ocean and land functions. Equation (15) is useful in that it allows the presence of coastlines to dominate the spectrum of $f(\Omega)$ at higher degrees (rather than short wavelength variability within the two spatial domains).

4 Application to Hydrological Loading

4.1 Physical Constraint: Sea Level Given the Total Load

We have applied the above concepts to the problem of partitioning the loading of water associated with the hydrological cycle between ocean and land. Here we apply a static oceanographic model to derive a set of physical constraints that can be used to

reduce the number of free parameters in the problem. Over the ocean the water load can be parameterized as relative sea level, and over land it can be considered as an equivalent thickness of water.

We assume:

1. redistribution of water on land is arbitrary;
2. the total mass of surface water is conserved, $f_{00} = 0$;
3. the sea surface is an equipotential, which generally does not equate to the (deformed) geoid;
4. the ocean bottom deforms with the solid Earth;
5. the solid Earth deforms in response to the spectral coefficients of the total load according to load Love number theory [*Farrell*, 1972].

The weakest assumption is number three, in that the ocean has dynamic topography due to ocean currents that may vary. For purposes of demonstrating the technique, however, ocean topography is assumed to be in static equilibrium. This is in any case instructive to compute the effect of land loading on relative sea level, which to first order might be considered uncoupled to ocean currents.

Under these conditions it can be shown [*Blewitt and Clarke*, 2003] that there is a linear relationship between the spectral coefficients of relative sea level and the total load for degrees other than zero:

$$\begin{aligned} \tilde{s}_{nm} &= \varsigma_n f_{nm} \qquad \text{for} \, n \geq 1 \\ \varsigma_n &\equiv \frac{3(1+k_n-h_n)\rho_S}{(2n+1)\rho_E} \end{aligned} \tag{17}$$

where $\rho_E = 5514\,\text{kg m}^{-3}$ is the mean density of the Earth, $\rho_S = 1025\,\text{kg m}^{-3}$ is the density of seawater; the degree-n load Love numbers are k_n for the gravitational potential and h_n for surface height of the solid Earth (not to be confused with tidal Love numbers). The first term $1+k_n$ in the numerator of equation (17) relates to the sea surface taking on the shape of the deformed geoid, and the last term $-h_n$ accounts for the change in relative height of the sea surface above the deformed ocean basin.

In contrast, the degree-0 term of relative sea level is not governed by the degree-0 term of the total load, which vanishes by conservation of mass $f_{00} = 0$, but rather relates to the balance between the volume of the ocean basin enclosed by the deformed geoid, and the water available to fill that basin:

$$\tilde{s}_{00} = \sum_{n',m'}^{n} [1-\varsigma_{n'}] \frac{c_{n'm'} f_{n'm'}}{c_{00}} \tag{18}$$

where the first term in the square bracket is related to change in global mean sea level from mass exchange with the land, and the smaller second term [*Dahlen,* 1976] is the contribution of the geoid-capped basin volume, which is generally non-zero even in the absence of mass exchange.

Equations (17) and (18) may be combined and expressed in matrix form:

$$\tilde{\mathbf{s}} = \mathbf{B}\mathbf{f} \tag{19}$$

where $\mathbf{B}$ is a square $(n \times n)$ matrix. Thus relative sea level can be computed up to the same degree and order as for the total load.

4.2 Total Load Given the Land Load

The total load is related to the smooth land load coefficients by application of the physical constraints given by equation (19) to the general partitioned spectral model, equation (15):

$$\begin{aligned} \mathbf{f} &= \mathbf{C}\mathbf{B}\mathbf{f} + (\mathbf{I}-\mathbf{C})\tilde{\mathbf{l}} \\ (\mathbf{I}-\mathbf{C}\mathbf{B})\mathbf{f} &= (\mathbf{I}-\mathbf{C})\tilde{\mathbf{l}} \end{aligned} \tag{20}$$

Given a spectral model of the land load $\tilde{\mathbf{l}}$ to degree $\bar{n}$ we can invert for the spectral model of the total load to any desired degree n

$$\mathbf{f} = (\mathbf{I}-\mathbf{C}\mathbf{B})^{-1}(\mathbf{I}-\mathbf{C})\tilde{\mathbf{l}} \tag{21}$$

which is well-conditioned, due fundamentally to the small ratio of the density of water to that of the Earth. This specific application of Clebsch-Gordan coefficients might be used, for example, to compute the extra contribution of the ocean to predictions of geocenter motion or polar motion from continental hydrology. In effect, the ocean acts to amplify the signal caused by land loading through equation (21).

4.3 Partitioned Load Given the Total Load

Alternatively, we might already have an estimate of the spectral coefficients of the change in total surface load based on the geodetically measured change in the Earth's shape or gravity field. In this case, the spectral model of the partitioned load is computed in two steps by equations (19) and (15):

$$\begin{aligned} \tilde{\mathbf{s}} &= \mathbf{B}\mathbf{f} \\ \tilde{\mathbf{l}} &= (\mathbf{I}-\mathbf{C})^{-1}(\mathbf{f}-\mathbf{C}\tilde{\mathbf{s}}) \end{aligned} \tag{22}$$

The maximum degree of partitioned models $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{l}}$ is that of the given total load **f**.

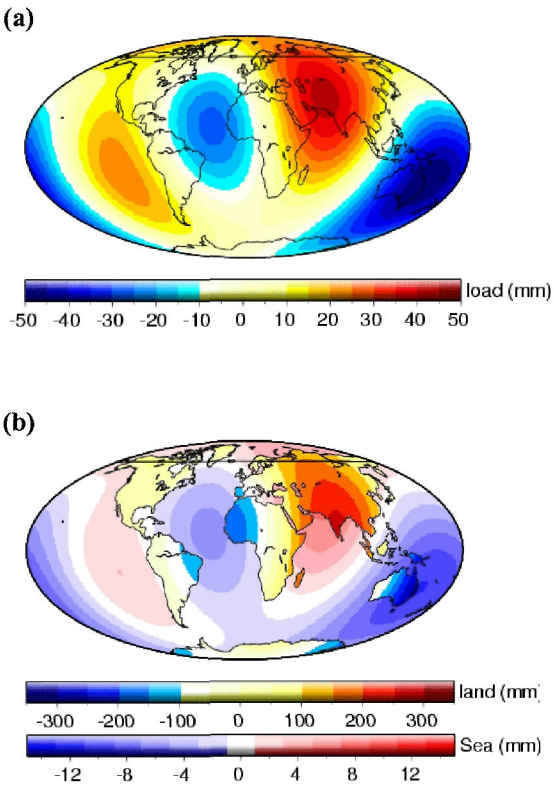


Fig. 1 Example of spatial partitioning of a global load function: **(a)** naïve spatial representation using a simple truncated spherical harmonic expansion; **(b)** partitioned function, applying physical constraints on static ocean topography.

4.4 Demonstration using Real Data

Figure 1 shows an example of our solution using GPS data for partitioned ocean-land loads given an initial naïve model, equation (12) up to degree and order 2. As a rule, we find that the ocean load follows a spatially similar pattern as the land load, except that it is reduced an order of magnitude. Not applying the partitioning method might therefore greatly overestimate the ocean load near coastlines.

5 Isomorphic Frames

The previous development assumes that we already have an estimate of the spectral coefficients of the total surface load. Such estimates can derive from GPS measurements of variation in Earth's shape, in an appropriate reference frame that is consistent with the degree-1 load Love numbers. Here we highlight the theoretical development and findings of *Blewitt* [2003] and *Blewitt and Clarke* [2003].

A vector surface displacement function on a sphere can be expressed [*Grafarend*, 1986]:

$$\mathbf{D}(\Omega) = \nabla \Psi(\Omega) + \nabla \times \left(\Gamma(\Omega)\hat{\mathbf{r}}\right) + H(\Omega)\hat{\mathbf{r}} \qquad (23)$$

where $H(\Omega)$ is the height function, $\Psi(\Omega)$ is the "poloidal" function, $\Gamma(\Omega)$ is the "toroidal" function, and ∇ is the surface gradient operator. For normal loading problems, the toroidal component is zero. Expanding (23) in terms of spherical harmonics, we find the degree-1 term for the displacement function can be written as a vector equation:

$$\mathbf{D}_1(\Omega) = \boldsymbol{\Psi}_1 + \hat{\mathbf{r}}\left[\left(\mathbf{H}_1 - \boldsymbol{\Psi}_1\right).\hat{\mathbf{r}}\right] \qquad (24)$$

where we define the spatially constant vectors $\boldsymbol{\Psi}_1 = (\Psi_{11}^C, \Psi_{11}^S, \Psi_{10}^C)$ and $\mathbf{H}_1 = (H_{11}^C, H_{11}^S, H_{10}^C)$, using notation for classical (real) spherical harmonics. The same degree-one deformation, as observed in a reference frame that has an origin translated by vector $-\Delta\mathbf{r}_0$ with respect to the original frame, will have (primed) surface displacements:

$$\begin{aligned} \mathbf{D}_1'(\Omega) &= \mathbf{D}_1(\Omega) + \Delta\mathbf{r}_0 \\ &= \boldsymbol{\Psi}_1' + \hat{\mathbf{r}}\left[\left(\mathbf{H}_1' - \boldsymbol{\Psi}_1'\right).\hat{\mathbf{r}}\right] \end{aligned} \qquad (25)$$

where

$$\begin{aligned} \boldsymbol{\Psi}_1' &= \boldsymbol{\Psi}_1 + \Delta\mathbf{r}_0 \\ \mathbf{H}_1' &= \mathbf{H}_1 + \Delta\mathbf{r}_0 \end{aligned} \qquad (26)$$

This implies a perhaps surprising result: there exist two special reference frames in which either the horizontal displacements or the vertical displacements are zero for degree-1 deformation [*Blewitt et al.*, 2001]. These are the CL (center of lateral figure) and CH (center of height figure) frames, respectively [*Blewitt*, 2003]. That the vertical degree-1 displacements can be made zero by a simple translation proves that the model Earth retains a shape of constant radius under a degree-1 deformation, and hence retains a perfect spherical shape. However, if $\mathbf{H}_1' = 0$, in general $\boldsymbol{\Psi}_1' = \boldsymbol{\Psi}_1 - \mathbf{H}_1 \neq 0$; so the surface is strained, as detected by GPS [*Blewitt et al.*, 2001] and by very long baseline interferometry [*Lavallée and Blewitt*, 2002].

Degree-1 displacement vector functions (including the degree-1 geoid height vector $\mathbf{N}_1$) relate to degree-1 load Love numbers [*Blewitt et al.*,

2001]:

$$(\mathbf{H}_1, \mathbf{\Psi}_1, \mathbf{N}_1) = -(h_1, l_1, 1+k_1)\Delta\mathbf{r}_{CE} \quad (27)$$

where $\Delta\mathbf{r}_{CE}$ is the "geocenter" displacement of the solid Earth center of mass CE relative to the stable center of mass of the entire Earth system CM:

$$\Delta\mathbf{r}_{CE} = -\frac{\rho_S}{\rho_E}\left(f_{11}^C \; f_{11}^S \; f_{10}^C\right)^T \quad (28)$$

Degree-1 load Love numbers themselves must be transformed if equation (27) is to be retained under a frame transformation, which requires that $\Delta\mathbf{r}_0$ be aligned with the geocenter displacement:

$$\Delta\mathbf{r}_0 = \alpha\Delta\mathbf{r}_{CE} \quad (29)$$

so that from equations (26), (27) and (29):

$$\begin{aligned}(\mathbf{H}_1', \mathbf{\Psi}_1', \mathbf{N}_1') &= -(h_1', l_1', 1+k_1')\Delta\mathbf{r}_{CE} \\ &= -(h_1, l_1, 1+k_1)\Delta\mathbf{r}_{CE} + \Delta\mathbf{r}_0 \\ &= -(h_1, l_1, 1+k_1)(1-\alpha)\Delta\mathbf{r}_{CE}\end{aligned} \quad (30)$$

Hence we can conclude that degree-1 load Love numbers transform as:

$$(h_1', l_1', 1+k_1') = (h_1, l_1, 1+k_1)(1-\alpha) \quad (31)$$

We call such special frames "isomorphic frames", as theoretically these are the only reference frames in which the degree-1 loading equation takes on the form of equation (27). Fortunately several frames in current use are isomorphic, including the CH, CL, CM, CE, and CF frames [*Blewitt*, 2003]. Frames that are not isomorphic include those where individual station height velocities are constrained.

Most commonly realized in GPS is the CF frame, in which the average vector surface displacement is constrained to zero. However, modelers tend to present results in the CE frame. Equation (31) allows the degree-1 load Love numbers to be transformed from the CE to the CF frame. The isomorphic parameter can be calculated for specific frames where the origin is defined either geometrically (with respect to observed surface displacements) or physically.(with respect to center of mass) [*Blewitt*, 2003]. In this case the appropriate value for the isomorphic parameter $\alpha = -0.021$, which is a small correction due to the CE and CF frame being almost identical. However, much larger values are possible for other isomorphic transformations.

6 Conclusions

We have developed two methodological tools that are useful for imposing self-consistency when inverting geodetic data for changes in global surface loading. The first tool considers the surface load as the multiple of two functions, a problem that can be transformed into the spectral domain by using Clebsch-Gordan coefficients. The CG method is demonstrated by imposing physical constraints on static ocean topography onto an assumed spectral loading distribution, so as to produce a realistic spatial representation of the load. The CG method may find broader application in geodesy to any spherical function that can be expressed as the multiple of a signal function, and a second known function, such as a filter function or an a priori model function.

The second tool transforms degree-1 load Love numbers into isomorphic frames in common use. Frames that are not isomorphic do not obey Love number relations, and so should be avoided.

Acknowledgements. The research was funded in the USA by NSF Geophysics and NASA Solid Earth and Natural Hazards, and in the UK by NERC. Thanks to G. Balmino for ocean loading coefficients. IGS provided the GPS data.

References

Balmino, G., K. Lambeck, and W.M. Kaula, A spherical harmonic analysis of Earth's topography, *Journ. Geophys. Res., 78,* 478-481, 1973.

Blewitt, G. and P. Clarke, Inversion of Earth's changing shape to weigh sea level in static equilibrium with surface mass redistribution, *Journ. Geophys. Res., 108* (B6), 2311, doi:10.1029/2002JB002290, 2003.

Blewitt, G., Self-consistency in reference frames, geocenter definition, and surface loading of the solid Earth, *Journ. Geophys. Res.*, doi:10.1029/2002JB002082, 2003.

Blewitt, G., D. Lavallée, P. Clarke, and K. Nurutdinov, A new global mode of Earth deformation: Seasonal cycle detected, *Science, 294*, 2,342-2,345, 2001.

Dahlen, F.A., The passive influence of the oceans upon the rotation of the Earth, *Geophys. Journ. Roy. Astron. Soc., 46*, 363-406, 1976.

Farrell, W.E., Deformation of the Earth by surface loads, *Rev. Geophys. and Space Phys., 10,* 761-797, 1972.

Grafarend, E.W., Three-dimensional deformation analysis: global vector spherical harmonic and local finite element representation, *Tectonophysics, 130*, 337-359, 1986.

Lavallée, D., and G. Blewitt, Degree-1 Earth deformation from very long baseline interferometry, *Geophys. Res. Lett., 29*(20), 1967, doi:10.1029/2002GL015883, 2002.

Particle Data Group (Hagiwara K., et al.), 2002 review of particle physics, *Phys. Rev. D 66*, 010001, 2002.

Wahr, J.M., The effects of the atmosphere and oceans on the Earth's wobble. 1. Theory, *Geophys. Journ. Roy. Astron. Soc., 70*, 349-372, 1982.

Role of algebra in modern day Geodesy

Joseph L. Awange, Yoichi Fukuda and Shuzo Takemoto
Department of Geophysics, Kyoto University, Kitashirakawa Oiwake-cho, Kyoto-shi Sakyo-ku, 606-8502, Japan. Email: awange@kugi.kyoto-u.ac.jp, Tel:+81-75-753-3319

Erik W. Grafarend
Department of Geodesy and GeoInformatics, Stuttgart University, Geschwister-Scholl str.24D, 70174 Stuttgart, Germany.

Abstract. Algebra, in particular the "Abelian Group" and the "Semi Group" (also known as "Monoid") axioms, which form a "ring with identity", are employed to define the "polynomial ring". Polynomial ring theory enables the solution of geodetic observations that can be converted into (algebraic) polynomials. The advantages of algebraic approaches are that they provide exact solutions to problems requiring closed form approaches (e.g. solving for geocentric coordinates from Helmerts projection through minimum distance mapping) and also act as tools to control iterative procedures. As a motivation, we present several examples of geodetic problems solved algebraically. These examples include; nonlinear analysis of bending angles in GPS-Meteorology, transformation of geocentric Cartesian coordinates into ellipsoidal, densification problems etc. The overriding advantage of the algebraic approach is the removal of the requirement of approximate starting values; they are non-iterative and enable detection of outliers.

Keywords: Polynomial ring, Groebner basis, Multipolynomial resultant, densification, GPS-Meteorology

1 Introduction

Algebra has been widely applied in fields such as Robotics for kinematic modelling of robots, Engineering for offset surface construction in solid modelling, Computer Science for automated theorem proving, Computer Aided Design etc. Recent application of algebra in Geodesy is evidenced in the work of Lannes and Dur (2003) who propose a new approach to differential GPS based on algebraic graph theory. The present contribution expounds on the role that algebra, in particular the *polynomial ring theory*, could play in modern Geodesy to solve systems of nonlinear equations.

In Geodesy, Photogrammetry and Computer Vision, *nonlinear equations* are often encountered in several applications as they often relate the observations (measurements) to the unknown parameters to be determined. In case the number of observations n and the number of unknowns m are equal ($n = m$), the unknown parameters may be obtained by solving explicitly or in a closed form the system of equations relating observations to the unknown parameters. For example, Cox et al. (1998, pp.28-32) has illustrated that for systems of equations with exact solutions, the system become vulnerable to small errors introduced during root findings and in case of extending the partial solution to the complete solutions of the system, the errors may accumulate and thus become so large. If the partial solution was derived by iterative procedures, then the errors incurred during the root finding may blow up during the extension of the partial solution to the complete solutions (back substitution). In some applications, symbolic rather than numerical solution are desired. In such cases, explicit procedures are usually employed. The resulting symbolic expressions often consist of univariate polynomials relating the unknown parameters (unknown variables) to the known variables (observations). By inserting known values into these univariate polynomials, numerical solutions are readily computed for the unknown variables.

Advantages of explicit solutions have been listed by Merritt (1949) as; *provision of satisfaction to the users (Photogrammetrist and Mathematicians) of the methods, provision of data tools for checking the iterative methods, desired by Geodesist whose task of control densification does not favor iterative procedures, provision of solace* and *the requirement of explicit solutions rather than iterative by some applications*. In Geodesy for example, the *Minimum Distance Mapping* problem considered by Grafarend and Lohse (1991) relates a point on the Earth's topographical surface uniquely (one-to-one) to a point on the *International Reference Ellipsoid.*

The solution of such an optimization problem requires that the equations be solved explicitly.

In this contribution, the term polynomial ring is defined and some areas where it has been applied in Geodesy mentioned. A complete theoretical concept of polynomial ring and a detail treatment of its application in Geodesy have been presented in Awange et al. (submitted).

In Section 2 is presented a definition of *polynomial ring,* while section 3 outlines geodetic problems solved using algebra. An example of GPS-Meteorology is presented in the Appendix.

2 Polynomial ring

In order to completely understand the term polynomial ring, one needs to familiarize with terms such as monomials, polynomials and linear algebra which we refer to Awange et al. (submitted a) or any algebraic book such as Cox et al. (1998) and McCoy and Janusz (2001).

Definition 2-1. A *sub-algebra* is called *a ring with identity* if the following two conditions encompassing (seven conditions) hold:

(a) The set R is an *Abelian group* with respect to addition, i.e. four conditions $\{G1+,G2+,G3+,G4+\}$ of *Abelian group* that describe associativity, identity, inverse and commutativity with respect to "addition" hold.

(b) The set R is a *semi-group* (monoid) with respect to multiplication; that is, associativity and identity $\{G1*,G2*\}$ holds.

The last condition is the left and right additive distributivity with respect to internal multiplication $\{D1*+\}$, which connects the *Abelian group* and the *monoid*. The collections of these axioms in Figure 2-1 comprise a *ring*.

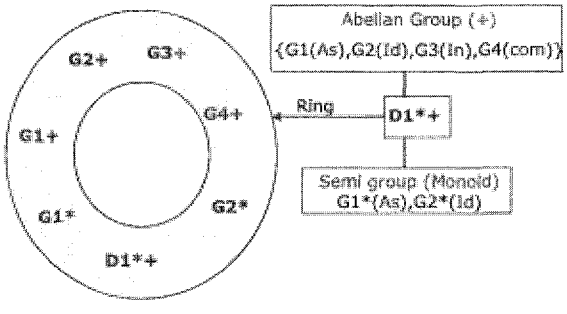


Fig 2-1.Ring

Definition 2-1 of a ring leads to the definition of the "*Polynomial ring*" as

Definition 2-2. Consider a ring R. If we consider an indeterminate $x \ddot{I} R$, assigning coefficients $a_i \hat{I} R$ to the indeterminate and obtaining the summation over finite number of distinct integers forms a *univariate polynomial*. Thus

$$f(x) = \mathop{\mathring{a}}_{a} a_a x^a, a_a \hat{I} R, a \,^3\, 0$$

is said to be a *univariate polynomial* over R. If two Polynomials be given such that $f_1(x) = \mathop{\mathring{a}}_{i} a_i x^i$ and $f_2(x) = \mathop{\mathring{a}}_{j} b_j x^j$, then two binary operation "*addition*" and "*multiplication*" can be defined on these polynomials such that:

(a) Addition:

$$f_1(x) + f_2(x) = \mathop{\mathring{a}}_{k} c_k x^k, c_k = a_k + b_k, c_k \hat{I} R$$

(b) Multiplicative:

$$f_1(x).f_2(x) = \mathop{\mathring{a}}_{k} c_k x^k, c_k = \mathop{\mathring{a}}_{i+j} a_i b_j, c_k \hat{I} R.$$

A collection of polynomials with these "*additive*" and "*multiplicative*" rules forms a polynomial ring. Designated P *polynomial ring* is represented by n unknown variables x over k and is expressed as $P := k[x_1,...,x_n]$. Its elements are polynomials known as *univariate* when $n = 1$ and *multivariate* otherwise. In order to be able to solve nonlinear systems of equations in Geodesy, they have to be expressed as elements of polynomial ring defined above using the *Theorem* 2-1 given in Awange and Grafarend (2003a) and solved using *Groebner basis* or *Multipolynomial resultant* approaches discussed in Awange and Grafarend (2003b,c).

3 Applications

In what follows, only outline of geodetic problems solved using polynomial ring algebra are presented. References are given where the problems have been treated in details.

3.1 GPS-Meteorology

In GPS-Meteorology, the determination of the refraction angle a (Appendix, Figure A-2) from

the measured excess phase is the beginning of the computational process to retrieve the atmospheric profiles of *temperature, pressure, water vapour* and *geopotential heights* (see Appendix for algebraic approach). For occultation CH-AI-2-PD+2002-123-13-0133-gold-002, it is established that the error of neglecting nonlinearity in computing bending angles using iterative approach that assumes nonlinearity increases with decreasing atmospheric height (Figure 3-1) leading to the deviation in computed temperature profile as in Figure 3-2.

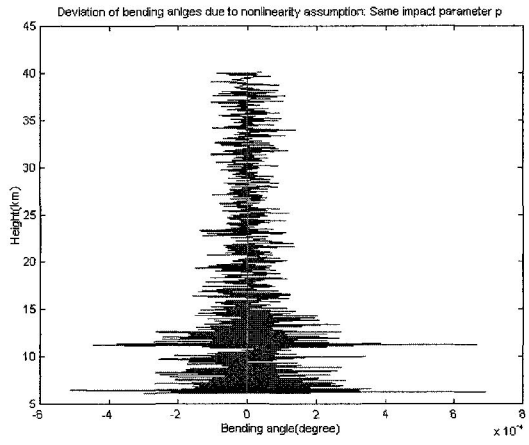


Fig. 3-1. Bending angle errors due to nonlinearity assumption using iterative procedures.

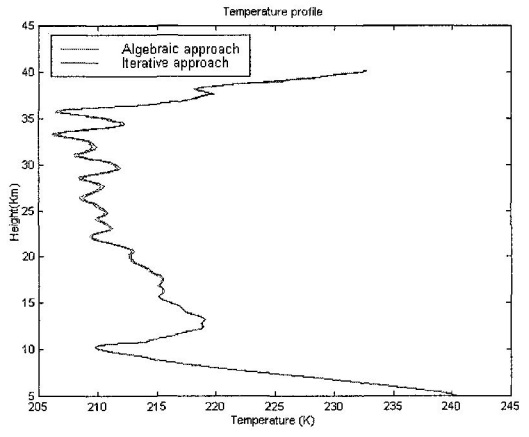


Fig. 3-2. Temperature profiles from algebraic and iterative approaches.

3.2 Topography to Ellipsoidal Mapping

In order to relate a point P on the Earth's topographic surface to a point p on the *International Reference* Ellipsoid} $E^2_{a,a,b}$ (Figure 3-3), one works with a bundle of half straight lines called *projection lines* which depart from P and intersect $E^2_{a,a,b}$ either not at all or in two points. There is *one projection line* which is at *minimum distance* relating P to p. Such an *optimization problem* is formulated by means of the Lagrangean $\pounds(x_1,x_2,x_3,x_4)$, which can be solved algebraically. An example based on Baltic Sea Level Project (Figure 3-4) has been treated in Awange et al. (submitted)

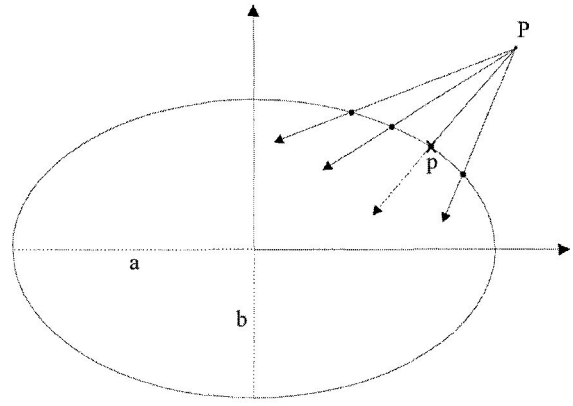


Fig. 3-3. Minimum distance mapping of a point P on the Earth's topographic surface to a point p on the *International Reference Ellipsoid* $E^2_{a,a,b}$.

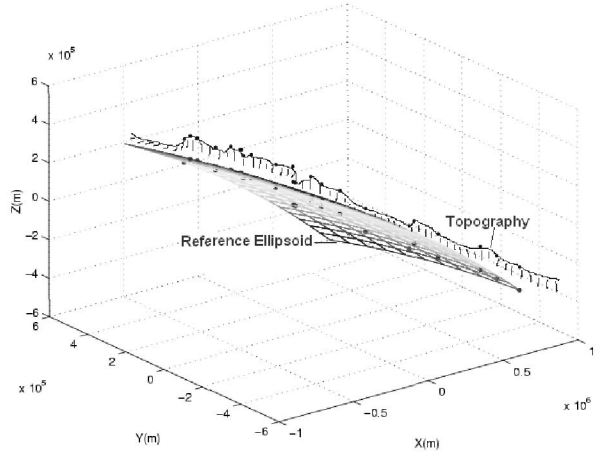


Fig. 3-4. Baltic Sea Level Project topographic points mapped on to the *International Reference Ellipsoid* $E^2_{a,a,b}$.

3.3 Global Positioning System GPS

Positioning with GPS requires the solution of four nonlinear observation equations of type pseudo-range or phase in what is known as four point problem in GPS positioning as illustrated in Figure (3-5). Awange and Grafarend (2002a,b) solve the nonlinear system of equation for both closed and overdetermined problems respectively using algebra and provide an algebraic tool for solving the initial receiver fix (e.g. http://www.ngs.noaa.gov/gps-toolbox/awange.htm).

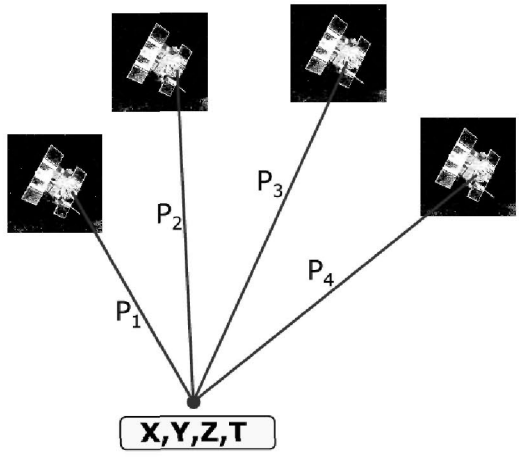


Fig. 3-5. GPS Point positioning using GPS satellites.

3.4 Geodetic network densification

Geodetic network densifications (e.g. Figure 3-6) are always carried out using three-dimensional ranging, intersection or resection. Densification problem becomes vital in areas where GPS receivers fail such as in tunnels and forest areas illustrated in Figure (3-6). Using algebra, Awange et al. (2003a,b) solve in closed form and overdetermined respectively the ranging problem. The point densification by resection and intersection is solved in Awange and Grafarend (2003a,b,c) and Awange et al. (2003c) respectively.

3.5 7-parameter datum transformation

The seven parameters transformation $C_7(3)$ has been solved in closed and overdetermined forms using algebra in Awange et al. (in press) and Awange and Grafarend (2003d) respectively. The scale parameter is given by a *quartic univariate polynomial* while expressions for translation and rotation are linear in scale parameter.

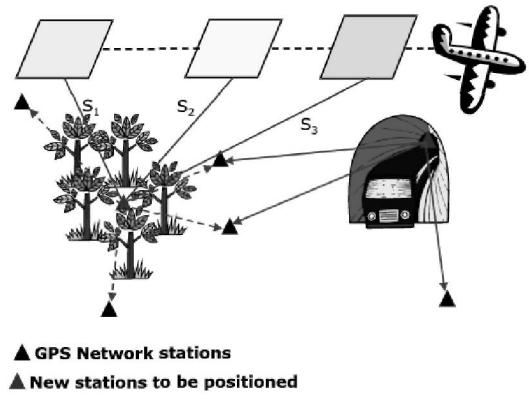


Fig. 3-6. Point densification in forest areas or inside a tunnel.

3.6 Industrial application

In the recently developed *Spatial Reference System* (Figure 3-7) that is designed to check and control the accuracy of the three-dimensional coordinate measuring machines and tooling equipment (Jurisch et al. 2003), the coordinates of the edges of the instrument are computed from distances of the bars. The nonlinear distance equations could be solved algebraically as already discussed in Awange et al. (2003a,b) to obtain the edge coordinates.

Fig 3-7. Spatial Reference System (SRS)

Appendix: Example of application of algebra to GPS –Meteorology.

Radio occultation with GPS takes place when a LEO satellite as illustrated in Figure A-1 views a transmitting GPS satellite, setting or rising behind the Earth's limb. GPS satellites send navigation signals, which passes through successively deeper layer of the Earths atmosphere and are received by LEO satellites. These signals are bend and retarded causing a delay in the arrival at the LEO satellite.

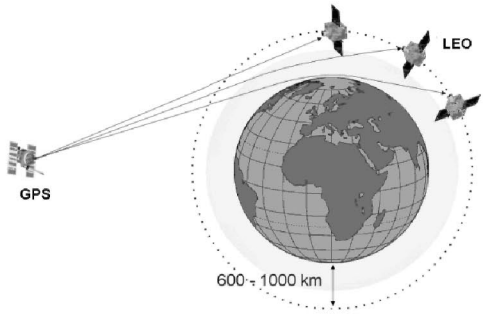


Fig. A-1. GPS Radio occultation .

Figure A-2 indicates the occultation geometry where the signal is send from GPS to the LEO satellite passing through the dispersive ionosphere and the atmosphere. As the signal is bend, the total bending angle α , an impact parameter a , and a tangent radius r are the desired parameters. The refraction angle α is accurately measured and related to atmospheric parameters of *temperature*, *pressure* and *water vapour* via the refractive index. Use is made of radio waves where the LEO receiver measures, at the required sampling rate, the dual band *carrier phase* measurements, the *C/A and P-code group delay* measurements and the *signal strength* measurements made by the flight receiver (Wickert, 2002). The data is then processed to remove the errors arising from short time oscillator instabilities in the satellite and the receivers using at least one ground station and at least one satellite that is not being occulted. Once the observations have been corrected for possible sources of errors, the resulting *Doppler shift* is used to determine the refraction angle α (Figure A-2). The system of *nonlinear* trigonometric equations often solved iteratively for the bending refraction angles α comprise two equations given as

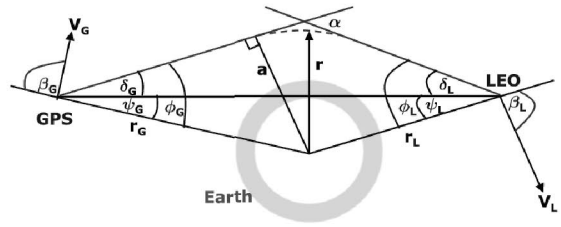


Fig. A-2: Radio Occultation geometry

$$v_l \cos(b_L - f_L) - v_G \cos(f_G + b_G) = a$$
$$r_G \sin f_G = r_L \sin f_L \qquad \text{(A1)}$$

with $\frac{dL_l}{dt}$ being the Doppler shift,

$$a = \frac{dL_l}{dt} + v_l \cos(b_L - y_L) - v_G \cos(y_G + b_G),$$

$\{v_L, v_G\}$ the projected LEO and GPS satellite velocities in the occultation plane and $\{r_L, r_G\}$ the radius of tangent points at LEO and GPS respectively. The angles are shown in Figure (A-2). Denoting $x = \sin f_G$, $y = \sin f_L$, $a_1 = v_L \cos b_L$, $a_5 = r_G$, $a_2 = v_L \sin b_L$, $a_3 = -v_G \cos b_G$, $a_6 = -r_L$, $a_4 = v_G \sin b_G$, where the signs of the velocities change depending on the directions of the satellites, and using Theorem 2-1 of Awange and Grafarend (2003a) together with the trigonometric addition formulae in (A1) leads to

$$a_1 \cos f_L + a_2 y + a_3 \cos f_G + a_4 x = a$$
$$a_5 x + a_6 y = 0 \qquad \text{(A2)}$$

In order to eliminate the trigonometric terms $\cos f_L$ and $\cos f_G$ appearing in (A2), they are taken to the right hand side and the resulting expression squared as

$$(a_2 y + a_4 x - a)^2 = (-a_1 \cos f_L - a_3 \cos f_G)^2 \qquad \text{(A3)}$$

which results in an expression having only trigonometric product $2a_1 a_3 \cos f_L \cos f_G$ on the right hand side once the squared trigonometric values $\cos^2 f_L$ and $\cos^2 f_G$ have been replaced by variables y, x and applying the trigonometric Pythagorean theorem of a unit circle. Arranging the resulting expression so as to have only $2a_1 a_3 \cos f_L \cos f_G$ on the right hand side and squaring both sides of the re-arranged expression and replacing the resulting squared trigonometric values $\cos^2 f_L$ and $\cos^2 f_G$ completes the conversion of (A1) into algebraic form

$$d_1x^4 + d_2x^3 + d_3x^3y + d_4x^2 + d_5x^2y^2 + d_6x^2y + d_7x + d_8xy^3 + d_9xy^2 + d_{10}xy + d_{00} = 0 \quad \text{(A4)}$$
$$a_5x + a_6y = 0$$

and $d_{00} = d_{11}y^4 + d_{12}y^3 + d_{13}y^2 + d_{14}y + d_{14}$. From (A4), one notes therefore that the solution of the nonlinear bending angle equation is actually obtained from the intersection of a fourth order polynomial and a line, indicating that there exist four solutions to the nonlinear problem.

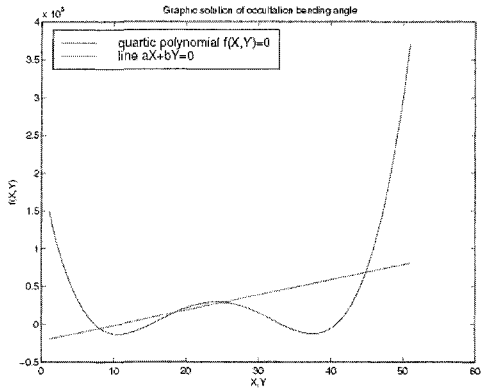


Fig. A-3 Algebraic curve for equation (4).

Equation (A4), an element of polynomial ring can be solved in exact form using the *Sylvester resultant* or *Groebner basis* algebraic approaches Awange and Grafarend (2003b,c). The differences in solutions of a in (A4) from both iterative and algebraic approaches are as plotted in Figure (3-3). Theory on GPS atmospheric sounding is presented in the works of Steiner (1998) and Wickert (2002).

Acknowledgment. The first author wishes to acknowledge the support of JSPS (Japan Society of Promotion of Science) for the financial support that enabled the undertaking of the study in Japan. The author is further grateful to his hosts and co-authors Professors S. Takemoto and Y. Fukuda for the good working atmosphere.

References

Awange JL, Grafarend E (2002a) Algebraic solution of GPS pseudo-ranging equations. Journal of GPS Solutions 4(5): 20-32

Awange JL, Grafarend E (2002b) Nonlinear adjustment of GPS observations of type pseudo-range. Journal of GPS Solutions 4(5): 80-93

Awange JL, Grafarend E (2003a) Explicit solution of the overdetermined three-dimensional resection problem. Journal of Geodesy 76: 605-616

Awange JL, Grafarend E (2003b) Groebner basis solution of the three-dimensional resection problem (P4P). Journal of Geodesy 77: 327-337

Awange JL, Grafarend E (2003c) Multipolynomial resultant solution of the three-dimensional resection problem (P4P). Bollettino di Geodesia e Science Affini 62, No 2

Awange JL, Grafarend E (2003d) Closed form solution of the overdetermined nonlinear 7-parameter datum transformation. Allgemeine Vermessungs Nachrichtern. 109: 130-148

Awange JL, Grafarend E, Fukuda Y, Takemoto S (2003a) Direct Polynomial approach to nonlinear distance (ranging) problems. Earth, Planets and Space 55: 231-241

Awange JL, Grafarend E, Fukuda Y, Takemoto S (2003b) Ranging algebraically with more observations than unknowns. Earth, Planets and Space

Awange JL, Grafarend E, Fukuda Y (2003c) Closed form solution of the triple three-dimensional intersection problem. Zeitschrift fuer Vermessungswessen 128(6)

Awange JL, Grafarend E, Fukuda Y (in press) Exact solution of the nonlinear 7-parameter datum transformation by Groebner basis. Bollettino di Geodesia e Science Affini

Awange JL, Grafarend E, Fukuda Y, Takemoto S (submitted) Polynomial ring algebra and its application to modern Geodesy. Journal of Geodesy

Cox D, Little J, O'Shea D (1998) Using algebraic geometry. Graduate Text in Mathematics 185. Springer-Verlag, New York.

Grafarend E, Lohse P, (1991) The minimal distance mapping of the topographic surface onto the (reference) ellipsoid of revolution. Manuscripta Geodaetica 16: 92-110

Jurisch, R, GG (2003) Kampmann and I. Schmadel, The Spatial Reference System (SRS)- an industrial application for geometric quality control introducing modern optimization, regulation and adjustment techniques. International Dimensional Workshop, Nashville Marriott Nashville, Tennessee, U.S.A. May 12 - 16, 2003.

Lannes A, Dur S (2003) Dual algebraic formulation of differential GPS. Journal of Geodesy 77: 22-29

McCoy NH, Janusz GJ (2001) Introduction to abstract algebra. Harcout Academic Press, San Diego

Merritt EL (1949) Explicit three-point resection in space. Phot. Eng. 15: 649-665

Steiner AK (1998) High resolution sounding of key climate variables using the radio occultation technique. Dissertation, Institute for Meteorology and Geophysics, University of Graz, No. 3

Wickert, J (2002) Das CHAMP- Radiookkultations-experiment: Algorithmen, Prozessierungssystem und Ergebnisse. Scientific Technical Report STR02/07, GFZ, Potsdam.

SYMPOSIUM G05

Geodynamics

C.R. Wilson

Mixed-mode GPS deformation monitoring – A cost-effective and accurate alternative?

Volker Janssen, Chris Rizos
School of Surveying and Spatial Information Systems, The University of New South Wales,
Sydney NSW 2052, Australia, email: v.janssen@student.unsw.edu.au

Abstract. During the past few years a methodology has been developed for processing data collected by GPS networks consisting of a mixed set of single-frequency and dual-frequency receivers. The strategy is to deploy a few permanent GPS stations with dual-frequency, geodetic-quality receivers surrounding an 'inner' deformation monitoring network of low-cost, single-frequency GPS receivers. The dual-frequency GPS network is used to generate a file of 'corrections', analogous to Wide Area DGPS correction models for the distance dependent biases. These 'corrections' are then applied to the double-differenced phase observations from the inner receivers to enhance baseline accuracy, primarily through empirical modelling of the residual atmospheric biases that otherwise would be neglected. Moreover, epoch-by-epoch baseline solutions are preferred in order to detect deformational signals in (near) real-time. Data from two continuous GPS networks have been used to investigate the performance of this configuration under severe ionospheric conditions and in different geographical regions. In the mid-latitude region the L1 baseline repeatability has clearly been improved by 40-50%, while an improvement of about 20% has been achieved in the equatorial region. The findings also indicate that the proposed procedure is sensitive to extreme ionospheric conditions, such as those experienced in close proximity to the geomagnetic equator during solar cycle maximum periods.

Keywords: Mixed-mode GPS, deformation monitoring, ionospheric corrections.

1 Introduction

The Global Positioning System (GPS) can be utilised in a wide range of deformation monitoring applications. The decreasing cost of GPS hardware, together with the increased reliability of the technology, facilitates such demanding applications as the monitoring of active volcanoes, tectonic fault lines, landslides, ground subsidence, bridges, dams, high-rise buildings, etc. GPS measurements can be continuous, automatic, conducted in all weather conditions, and provide three-dimensional positioning results.

Deformation monitoring using GPS is usually carried out by installing and operating a local network of GPS receivers located on and around the deforming body. Most continuous GPS deformation monitoring networks are comparatively costly because they rely entirely on the use of dual-frequency instrumentation (e.g. Ogaja *et al.*, 2001; Owen *et al.*, 2000; Wong *et al.*, 2001). In order to keep the cost of such a deformation monitoring system to a minimum, single-frequency GPS receivers need to be used. On the other hand, some atmospheric biases (mainly the ionospheric delay) cannot be accounted for directly if only one frequency is used, resulting in lower accuracy. It is usually assumed that the differential ionospheric and tropospheric delays between two receivers can be ignored if baseline lengths are kept below 10km. However, especially during solar sunspot cycle maximum periods, ionospheric disturbances have indeed corrupted L1 baseline measurements over distances less than 10km, adversely affecting baseline repeatability (Janssen *et al.*, 2001).

It is therefore necessary to combine a single-frequency deformation monitoring network with a small number of dual-frequency receivers surrounding the deformation zone in order to account for these atmospheric effects. Such a mixed-mode configuration offers considerable flexibility and cost savings for deformation monitoring applications, which require a dense spatial coverage of GPS stations, and where it is not possible, nor appropriate, to establish permanent GPS networks using dual-frequency instrumentation. Hartinger & Brunner (2000) have used such a mixed-mode approach to monitor landslides in the Austrian Alps.

2 Mixed-mode GPS deformation monitoring

In a mixed-mode GPS network approach, a fiducial network of dual-frequency receivers surrounding the deformation zone can be used to generate empirical 'correction terms'. These can then be applied to the single-frequency observations of the inner network on the deforming body to account for the ionospheric effects, and hence improve baseline accuracy. Figure 1 shows the ideal configuration of such a mixed-mode GPS network where the triangles denote fiducial stations, while the dots indicate single-frequency sites. The reference stations are to be situated outside the deformation zone but within the local tectonic region in order to avoid unwanted displacements of the external network.

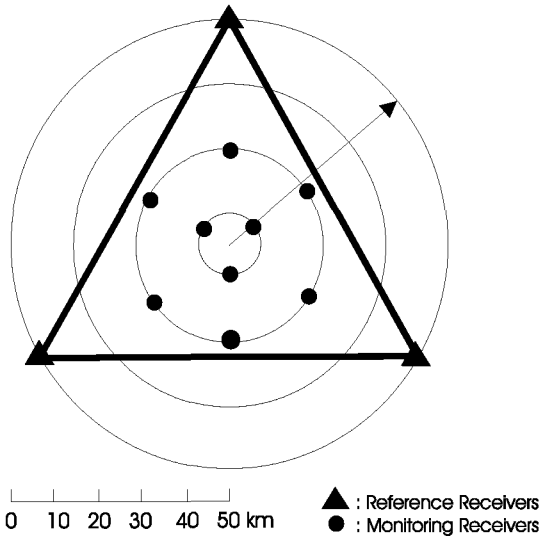


Fig. 1. Ideal network configuration of a mixed-mode GPS deformation monitoring network

A linear combination model, utilising a single ionospheric layer model at a height of 350km, is used to determine the corrections (Han & Rizos, 1996; Han, 1997). This approach can account for orbit bias and ionospheric delay, as well as somewhat mitigate tropospheric delay, multipath and measurement noise across the network. Data from the GPS reference station network can thus be used to derive empirical corrections to the double-differenced carrier phase data formed between the stations of the inner network.

If the number of GPS reference stations is assumed to be three, the double-differenced observation equation for a baseline between a reference station and a single-frequency user receiver located inside the triangle (denoted by u) can be expressed in the form of a linear combination (Han, 1997):

$$\nabla\Delta\phi_{u,3} - \left[\alpha_1 \cdot V_{1,3} + \alpha_2 \cdot V_{2,3}\right] = \nabla\Delta\rho_{u,3} + \lambda \cdot \nabla\Delta N_{u,3} + \varepsilon_{\nabla\Delta\phi} \quad (1)$$

where $\nabla\Delta$ is the double-differencing operator, ϕ is the carrier phase observation in units of metres, ρ is the distance between receiver station and satellite, N is the integer ambiguity for a particular satellite-receiver phase measurement, and $\varepsilon_{\nabla\Delta\phi}$ represents the carrier phase observation noise. The parameters α refer to the position of the user receiver inside the fiducial triangle and can be determined based on the following conditions (Han & Rizos, 1996; Wu, 1994):

$$\sum_{i=1}^{3}\alpha_i = 1,\ \sum_{i=1}^{3}\alpha_i \cdot \left(\vec{X}_u - \vec{X}_i\right) = 0 \text{ and } \sum_{i=1}^{3}\alpha_i^2 = \min \quad (2)$$

where $\vec{X}_u$ and $\vec{X}_i$ are the user station position vector and reference station position vector respectively, while α_i denotes the weight for reference station i.

The residual vectors are formed from the double-differenced observations between reference stations 1 & 3 and 2 & 3:

$$V_{1,3} = \nabla\Delta\phi_{1,3} - \nabla\Delta N_{1,3} - \nabla\Delta\rho_{1,3} \quad (3)$$

$$V_{2,3} = \nabla\Delta\phi_{2,3} - \nabla\Delta N_{2,3} - \nabla\Delta\rho_{2,3} \quad (4)$$

The 'correction term' $\left[\alpha_1 \cdot V_{1,3} + \alpha_2 \cdot V_{2,3}\right]$ can now be obtained and applied to the user receiver. For a baseline between two GPS stations j and k of the inner (single-frequency) network the linear combination can be expressed as:

$$\nabla\Delta\phi_{k,j} - \left[\alpha_1^{k,j} \cdot V_{1,3} + \alpha_2^{k,j} \cdot V_{2,3}\right] = \nabla\Delta\rho_{k,j} + \lambda \cdot \nabla\Delta N_{k,j} + \varepsilon_{k,j} \quad (5)$$

where $\alpha_i^{k,j}$ is the difference in the α_i value for stations j and k, and $\left[\alpha_1^{k,j} \cdot V_{1,3} + \alpha_2^{k,j} \cdot V_{2,3}\right]$ is the 'correction term' for the inner baseline between these user receivers.

By forming the double-differenced observables between the inner single-frequency receivers, and using the residual vectors generated from the reference stations, the inner stations' coordinates can be determined without the need to use any GPS reference station observations at all. In practice, holding one reference site fixed, the baselines to the other two sites are processed and 'correction terms' are obtained for both baselines. These are then scaled (by means of the parameters α_i) according to the position of the inner stations inside the reference network triangle to generate double-differenced corrections for the inner baselines.

3 GPS data processing

A single-frequency version of the Baseline software package developed at UNSW is used to process the inner network incorporating the 'correction terms', which are determined using a modified version of the Bernese software. For deformation monitoring applications multi-baseline processing strategies should be used because all baselines are then computed together, taking into account the between-baseline correlations which arise from observing a GPS network simultaneously. For continuous deformation monitoring a near real-time, epoch-by-epoch solution is preferred in order to detect movements over a short period of time.

4 Case study: Mid-latitude region

Data from the Southern California Integrated GPS Network (SCIGN) (SCIGN, 2003) were used to investigate the performance of the mixed-mode network configuration in the mid-latitude region. Figure 2 shows the location of the GPS sites, which are all equipped with dual-frequency receivers. The part of the network used in this study consists of an outer network of three

sites (FXHS, FMTP, QHTP) surrounding an inner network of three sites (CSN1, OAT2, CMP9). The outer sites were used as fiducial GPS reference stations, indicated by triangles in Figure 2, while the inner sites (indicated by circles) simulated single-frequency receiver stations (by ignoring the observations made on L2). The data were collected under solar maximum conditions on three consecutive days from August 8-10, 2000 (DOY 221-223).

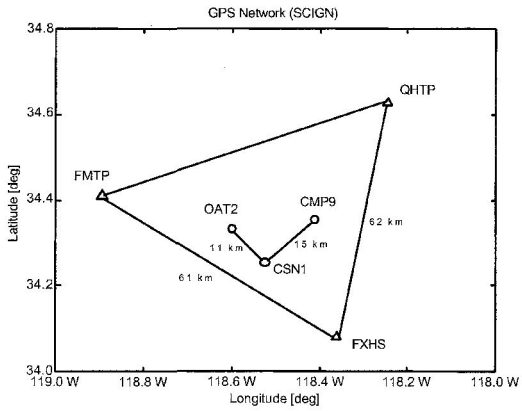


Fig. 2. SCIGN network stations used in this study

The fiducial baselines FXHS-FMTP and FXHS-QHTP are used to model the ionospheric conditions across the network. L1 'correction terms' for the inner baselines CSN1-OAT2 and CSN1-CMP9 can then be determined on a satellite-by-satellite and epoch-by-epoch basis, according to equation (5). As an example, Figure 3 shows the double-differenced corrections obtained over 24 hours for day 221. It can easily be seen that ionospheric activity in mid-latitude regions is mainly a daytime phenomenon. A comparison with directly generated ionospheric delay values for the same baselines (using dual-frequency data) showed that this procedure does indeed compute correct values for the inner baselines.

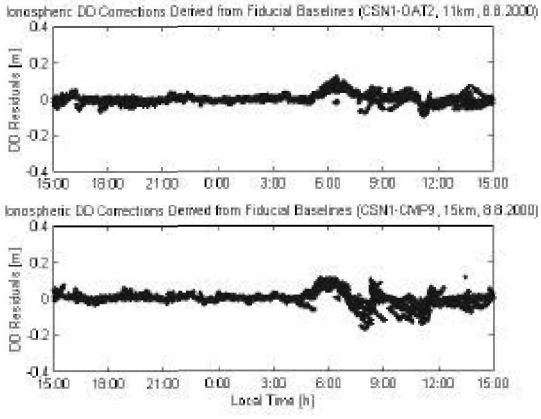


Fig. 3. Double-differenced L1 corrections for the inner baselines (DOY 221)

The Baseline software is then used to process the inner baselines in single-frequency mode, with and without applying the empirically derived corrections. It can readily be assumed that no ground deformation has taken place during the period of observation. Hence, the baseline repeatability gives a good indication of the accuracy that can be achieved with the data processing strategy described in this paper.

Figure 4 shows the results obtained for one inner baseline ignoring ionospheric effects, while Figure 5 shows the results obtained by applying ionospheric corrections on day 221. The graphs show the Easting, Northing and Height components over a 24-hour period, each dot representing a single-epoch solution. The respective standard deviations are also included. In both cases the Saastamoinen model was used to account for the tropospheric bias, as recommended by Mendes (1999).

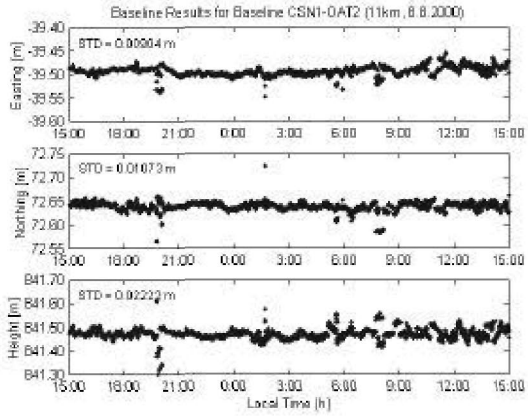


Fig. 4. Results for an inner baseline not using ionospheric corrections (DOY 221)

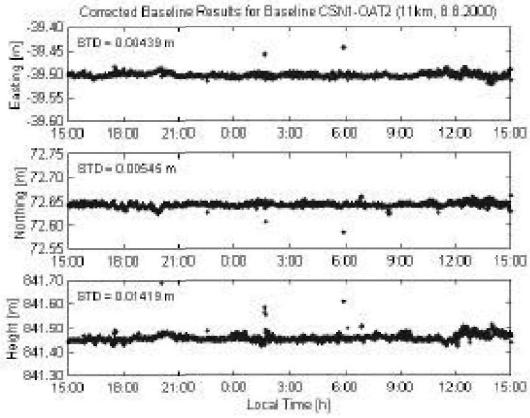


Fig. 5. Results for an inner baseline applying ionospheric corrections (DOY 221)

Similar results were obtained for the second inner baseline and on the following two days. Standard deviations of less than 1cm horizontally and 1.5-3cm vertically have been achieved for a single-epoch solution by applying the corrections. On average, the standard deviation of the baseline results has been reduced by almost 50% in the horizontal components and almost 40% in the vertical component (Table 1).

Table 1. Average improvement in the STD for both baselines on days 221-223

Baseline	DOY	East [%]	North [%]	Height [%]
CSN1-	221	51.4	49.2	36.1
OAT2	222	38.1	35.8	40.7
(11km)	223	53.8	51.4	33.6
CSN1-	221	54.7	59.5	48.4
CMP9	222	38.7	29.1	33.0
(15km)	223	56.1	55.9	34.1
Average [%]		49	47	38

5 Case study: Equatorial region

In order to investigate the performance of the mixed-mode network processing approach in the equatorial region, data from the Hong Kong Active GPS Network (Chen *et al.*, 2001) were used. Figure 6 shows the location of the GPS sites, which are all equipped with dual-frequency receivers. The network consists of an outer network of three sites (HKKY, HKFN, HKSL) surrounding two inner sites (HKKT, HKLT). Due to the absence of a third site within the fiducial triangle, the fiducial site HKKY was used to form the inner (single-frequency) baselines. Note that HKLT is located just outside the fiducial triangle, which should normally be avoided. However, this does not have any effect on the processing in this case, as HKKY-HKFN and HKKY-HKSL are used as fiducial baselines. The data were collected under solar maximum conditions on three consecutive days from October 11-13, 2000 (DOY 285-287).

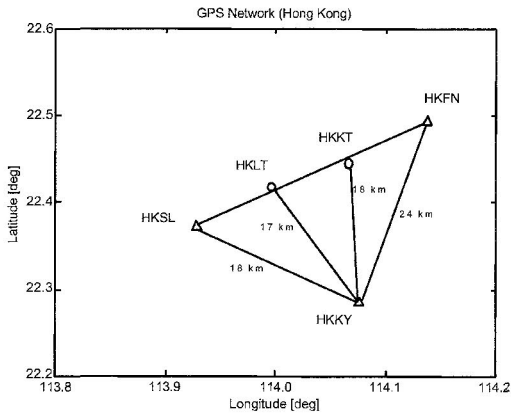


Fig. 6. Hong Kong GPS Active Network stations used in this study

According to equation (1) L1 'correction terms' can be determined for the inner baselines HKKY-HKKT and HKKY-HKLT. As an example, Figure 7 shows the double-differenced corrections obtained for the first day. A comparison with directly generated ionospheric delay values for the same baselines (using dual-frequency data) confirmed the correctness of these values. It is evident that ionospheric activity in the equatorial region is much more severe than in mid-latitudes, reaching values of a few cycles. As expected, most of the activity occurs after local sunset, although there is also a lot of daytime activity. This may be explained by intensified small-scale disturbances in the ionosphere during a period of increased solar activity. Furthermore, it can be identified as the primary diurnal maximum of the equatorial anomaly, also known as the *fountain effect* (high electron concentration observed on either side of the geomagnetic equator at magnetic latitudes of around 10-20°). The equatorial anomaly has been investigated by, amongst others, Huang & Cheng (1991) and Skone (2000).

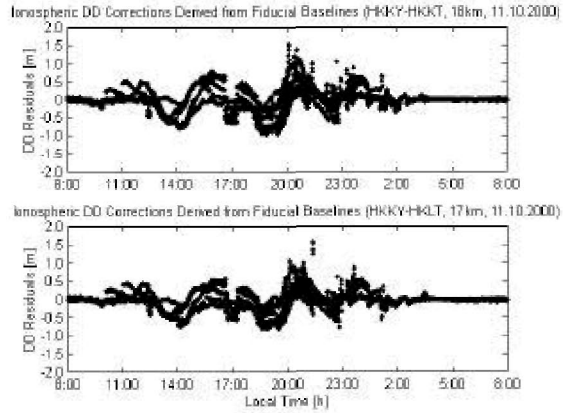


Fig. 7. Double-diff. L1 corrections for the inner baselines (DOY 285)

Again, the baseline processing results for the inner baselines obtained with and without applying the empirically derived corrections were compared. In both cases the Saastamoinen model was used to account for the tropospheric bias. As an example, Figure 8 shows the single-frequency results obtained for one inner baseline ignoring ionospheric effects, while Figure 9 shows the results obtained by applying ionospheric corrections on day 285. The graphs show the Easting, Northing and Height components over a 24-hour period, each dot representing a single-epoch solution. The respective standard deviations are also included. Similar results were obtained for the second inner baseline and on the following two days. Although not obvious from these figures, an improvement in accuracy of around 20% in all three components has been achieved by applying the corrections (as indicated in Table 2).

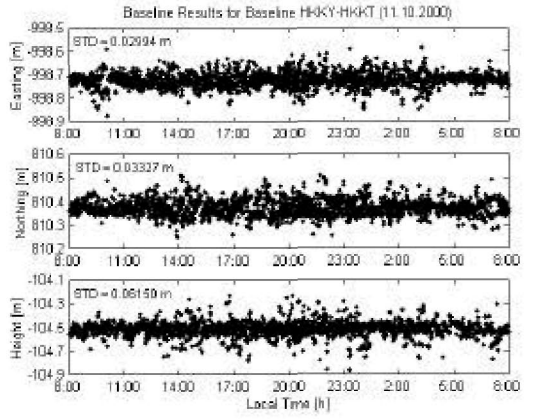


Fig. 8. Results for an inner baseline not using ionospheric corrections (DOY 285)

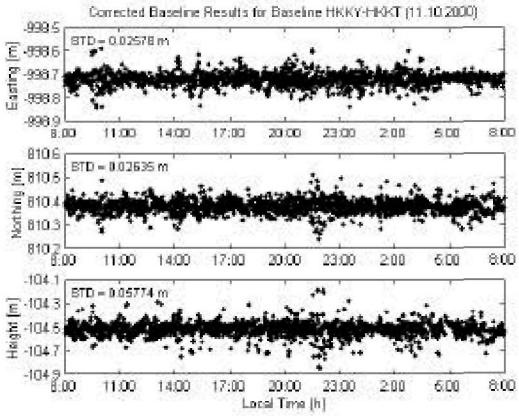


Fig. 9. Results for an inner baseline applying ionospheric corrections (DOY 285)

Table 2. Average improvement in the STD for both baselines on days 285-287

Baseline	DOY	East [%]	North [%]	Height [%]
HKKY-	285	13.9	20.8	6.1
HKKT	286	15.5	16.8	25.9
(18km)	287	23.6	24.9	26.9
HKKY-	285	15.2	19.0	5.9
HKLT	286	15.3	12.4	21.1
(17km)	287	25.4	26.4	26.3
Average [%]		18	20	19

However, in spite of the shorter fiducial baseline lengths compared to the SCIGN network, standard deviations could not be reduced below the few-cm level when corrections were applied. Values of 2.0-3.5cm in the horizontal components and 4.5-6.5cm in the height component are too high to permit reliable detection of ground deformation at the desired accuracy level. This emphasizes the significant effect of the ionosphere on GPS deformation monitoring networks, especially at low-latitudes in periods of heightened solar activity.

6 Concluding remarks

A procedure to process a mixed-mode GPS network for deformation monitoring applications has been described. Single-frequency GPS observations have been improved by generating empirical corrections obtained from a fiducial network of dual-frequency reference stations surrounding the inner single-frequency network. This method accounts for the ionospheric bias that otherwise would have been neglected when using single-frequency instrumentation only. Data from two GPS networks located in different geographical regions have been used to simulate such a network configuration in order to investigate the impact of the proposed processing strategy on baseline results.

In the mid-latitude region the single-frequency baseline repeatability has clearly been improved by applying the empirical 'correction terms'. The standard deviations of the baseline results have been reduced by 40-50% to less than 1cm horizontally and 1.5-3cm vertically for a single-epoch baseline solution.

In the equatorial region an improvement of approximately 20% in all three components has been achieved by applying the corrections. However, the standard deviations could not be reduced below the few-cm level. This indicates that extreme ionospheric conditions, such as those experienced in close proximity to the geomagnetic equator during solar cycle maximum periods, can reduce the efficiency of the proposed method. Nevertheless, the approach of processing a mixed-mode GPS network presented here can be a cost-effective and accurate tool for deformation monitoring suitable for a variety of applications. The correction generation algorithm could easily be implemented in a continuous GPS reference station network, distributing ionospheric corrections as a routine product.

Acknowledgements. SCIGN and its sponsors, the W.M. Keck Foundation, NASA, NSF, USGS and SCEC, as well as A/Prof. Peter Morgan from the University of Canberra, are thanked for providing the data used in this study. The first author is supported by an International Postgraduate Research Scholarship (IPRS) and funding from the Australian Research Council (ARC).

References

Chen W., Hu C., Chen Y., Ding X. and Kwok S.C.W. (2001): Rapid Static and Kinematic Positioning with Hong Kong GPS Active Network. Proc. ION GPS-2001, Salt Lake City, Utah, 11-14 September, 346-352.

Han S. (1997): Carrier Phase-Based Long-Range GPS Kinematic Positioning. PhD Dissertation, UNISURV S-49, School of Geomatic Engineering, The University of New South Wales, Sydney, Australia, 185pp.

Han S. and Rizos C. (1996): GPS Network Design and Error Mitigation for Real-Time Continuous Array Monitoring Systems. Proc. ION GPS-96, Kansas City, Missouri, 17-20 September, 1827-1836.

Hartinger H. and Brunner F.K. (2000): Development of a Monitoring System of Landslide Motions Using GPS. Proc. 9th FIG Int. Symp. on Deformation Measurements, Olsztyn, Poland, Sep 1999, 29-38.

Huang Y.-N. and Cheng K. (1991): Ionospheric Disturbances at the Equatorial Anomaly Crest Region during the March 1989 Magnatic Storms. Journal of Geophysical Research 96(A8): 13953-13965.

Janssen V., Roberts C., Rizos C. and Abidin H.Z. (2001): Experiences with a Mixed-Mode GPS-Based Volcano Monitoring System at Mt. Papandayan, Indonesia. Geomatics Research Australasia 74: 43-58.

Mendes V.B. (1999): Modeling the Neutral-Atmosphere Propagation Delay in Radiometric Space Techniques. PhD Dissertation, Dept. of Geodesy & Geomatics Eng. Tech. Rept. No. 199, University of New Brunswick, Fredericton, Canada.

Ogaja C., Rizos C., Wang J. and Brownjohn J. (2001): A Dynamic GPS System for On-line Structural Monitoring. Int. Symp. on Kinematic Systems in Geodesy, Geomatics & Navigation (KIS 2001), Banff, Canada, 5-8 June, 290-297.

Owen S., Segall P., Lisowski M., Miklius A., Murray M., Bevis M. and Foster J. (2000): January 30, 1997 Eruptive Event on Kilauea Volcano, Hawaii, as Monitored by Continuous GPS. Geophysical Research Letters 27(17): 2757-2760.

SCIGN (2003): Southern California Integrated GPS Network (SCIGN) website: http://www.scign.org/.

Skone S.H. (2000): Wide Area Ionosphere Modeling at Low Latitudes – Specifications and Limitations. Proc. ION GPS-2000, Salt Lake City, Utah, 19-22 September, 643-652.

Wong K.-Y., Man K.-L. and Chan W.-Y. (2001): Monitoring Hong Kong's Bridges – Real-Time Kinematic Spans the Gap. GPS World 12(7): 10-18.

Wu J.T. (1994): Weighted Differential GPS Method for Reducing Ephemeris Error. Manuscripta Geodaetica 20: 1-7.

Two examples of optimal design of geodynamic GPS networks

N.V. Shestakov, H.E. Waithaka, M. Kasahara
Institute of Seismology and Volcanology,
Hokkaido University, N10 W8, Kita-ku, Sapporo 060-0810, Japan
e-mails: nikon@phys.dvgu.ru; waithaka@eos.hokudai.ac.jp; mkasa@eos.hokudai.ac.jp

M. D. Gerasimenko
Institute of Applied Mathematics FEBRAS, Radio st., 7, Vladivostok 690041, RUSSIA
e-mail: mdger@iam.dvo.ru

Abstract. Many local geodynamic geodetic GPS networks are increasingly being exploited for monitoring crustal movements and deformations. It is thus very important that such monitoring networks should be developed and observed in the most economical way and at the same time satisfy some precision and reliability criteria set by the user. In this report we present our solution of the optimal design of two-dimensional geodynamic GPS networks which guarantees the required precision of parameters for the chosen deformation model with the least cost. A common cost criterion is used as a target function and the non-linear programming problem is solved for an optimal set of GPS vectors. The proposed methodology is illustrated using two optimization examples of working GPS networks using simple deformation models.

Keywords. Optimal design, geodynamic monitoring GPS network, non-linear programming task

1 Introduction

Today GPS networks are widely used for monitoring crustal movements and deformations. Accordingly, a reciprocal need for optimal design of monitoring networks is of great practical interest.

The problem of designing geodetic networks can generally be divided into four stages (Grafarend 1974). The Zero-Order Design (ZOD) deals with the definition of an optimum reference system of the network. The First-Order Design (FOD) involves the determination of geometric shape of the network (i.e. optimal number and location of geodetic stations). On the other hand, the Second-Order Design (SOD) deals with determination of the optimal weights of network measurements (which observations and with what precision should be made in the network). Lastly, the Third-Order Design (TOD) considers the improvement of an existing network (i.e. which additional measurements and with what precision or weights should be made with the objective of improving some properties of the network). In case we take into account the period of time between consecutive observations, the term fourth order design may be used (Schmitt 1982). Very often designing a real geodetic network is an integration of a few of these individual stages that may lead to a more complicated design process. The classical FOD problem has already been considered in previous works (Koch 1985; Schmitt 1982, 1985). Further, SOD examples of pure GPS networks are for instance explained in Even-Tzur and Papo (1996); Stopar (2001). Solution of the integrated FOD and SOD problems for traditional and GPS networks are considered in Afonin (1975); Kuang (1996). The solution of GPS vector configuration design problem based on sensitivity analysis of the geodetic network is given in Even-Tzur (2002).

In this report we apply the methodology proposed in Gerasimenko *et al.* (2000) for optimal design of two-dimensional deformation GPS networks. The task is to design a network project to provide the required precision of deformation parameters of a chosen deformation model with the least cost. Our results are based on the solution of the non-linear optimization problem using well-known mathematical methods.

2 Problem formalization

Any monitoring GPS network should be developed and observed using the most economical way and should satisfy some precision and reliability criteria set by the user. It is especially important in the case of large and dense geodynamic GPS net which contains a lot of observations and stations. In the case of campaign mode observations, field measurements have a significant contribution into campaign cost. If we consider the GPS baseline vector com-

ponents as the directly measured quantities, then the optimality criterion for economy can be written in the following form

$$\Phi = \sum_{i=1}^{3n(n-1)/2} c_i p_i \rightarrow \min, \quad (1)$$

where n is the total number of geodetic points and $3n(n-1)/2$ is the number of measured baseline components in a GPS network; p_i is the weight of observation i and c_i is the coefficient of relative profitableness of observation i. The most general and simplest way to calculate such coefficients is by using the expression

$$c_i = \frac{k_i}{k_1}, \quad (2)$$

where k_i is the expense related with the observation i and k_1 is the total expense of carrying out the first observation. In reality each coefficient k_i is comprised of many factors related with the cost of measurements in the field campaign, for example the cost of constructing and installing the geodetic monuments on observation points, driving to the sites, setting up the instruments and so on (Afonin 1975). It is possible to develop and use more complicated equations for calculation of coefficients c_i (for traditional geodetic observations see for example Provorov and Afonin 1976). As a result we may obtain different optimal projects for the same GPS network using different strategies for calculation of c_i.

The optimized weights of GPS observations should be within limits

$$0 \le p_i \le p_{i,\max}, \; i = 1, 2, ..., 3n(n-1)/2. \quad (3)$$

In other words optimal weights should be non-negative and limited by maximum achievable precision of corresponding observations provided by the chosen equipment, the post-processing strategy and so on. If some optimal weight p_i is equal or close to zero the corresponding observation does not contribute to the precision of deformation parameters and we can eliminate it from the optimal network scheme.

As we had earlier stated in the introduction the goal of our task is the optimal set of GPS observations to provide the required precision of the unknown parameters for a chosen deformation model of the investigated phenomenon (fault, volcano and so on). Such a model can be constructed as a function of displacements $\mathbf{d}$ of the network points (Chrzanowski *et al.* 1983) such that:

$$\mathbf{d} = \mathbf{X}_2 - \mathbf{X}_1 = \mathbf{B}\mathbf{e}, \quad (4)$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are vectors of coordinates of network points obtained from the post-processing of two different epochs of GPS observations; $\mathbf{B}$ is the design matrix of deformation model and $\mathbf{e}$ is the vector of unknown deformation parameters. For the construction of the design matrix $\mathbf{B}$ the approximate behavior of deformation parameters $\mathbf{e}$ as some mathematical model should be known a priori. Usually such information can be obtained from previous geological and/or geophysical investigations.

If the model errors are negligible compared with the errors of the displacements, the covariance matrix of the least-square estimate of $\mathbf{e}$ is given by the equation

$$\mathbf{Q}_\mathbf{e} = (\mathbf{B}^T \mathbf{Q}_\mathbf{d}^- \mathbf{B})^-, \quad (5)$$

where $\mathbf{Q}_\mathbf{d}$ is the covariance matrix of the vector of displacements $\mathbf{d}$. Assuming that the two measurement epochs are identical, then the covariance matrix $\mathbf{Q}_\mathbf{d}$ reduces to the form

$$\mathbf{Q}_\mathbf{d} = \mathbf{Q}_{\mathbf{X}_1} + \mathbf{Q}_{\mathbf{X}_2} = 2\mathbf{Q}_\mathbf{X} = \\ = 2(\mathbf{A}^T \mathbf{Q}^{-1} \mathbf{A})^- = 2\mathbf{N}^- \quad (6)$$

where $\mathbf{Q}_\mathbf{X}$ is the covariance matrix of coordinates of the network points; $\mathbf{Q}$ and $\mathbf{A}$ are the covariance matrix and the design matrix of GPS observations respectively; $\mathbf{N}^-$ denotes the generalized inverse of matrix $\mathbf{N}$. However forming the covariance matrix of GPS measurements is not so simple because all baseline components are correlated and affected by some common error sources, especially in the case of short baselines. But the introduction of erroneous or/and strong physical correlations in geodetic measurements (and GPS measurements as well) may lead to extraordinarily absurd and unexpected results as shown in Gerasimenko *et al.* (2002); Strang and Borre (1997). Taking this into account,

and for simplicity we will consider the baseline components as uncorrelated measurements in the next section. In this case $\mathbf{Q}$ will be a diagonal matrix.

The criteria of precision requirements for deformation parameters can be expressed as constraints on some functions of covariance matrix $\mathbf{Q}_{\mathbf{e}}$ (Gerasimenko *et al.* 2002) in the form

$$\varphi_i(\mathbf{Q}_{\mathbf{e}}) \le \varphi_{i,\max}(\mathbf{Q}_{\mathbf{e}}),\ i = 1, 2, ..., r, \qquad (7)$$

where $\varphi_i(\mathbf{Q}_{\mathbf{e}})$ are the chosen criteria of optimality and $\varphi_{i,\max}(\mathbf{Q}_{\mathbf{e}})$ are the maximum permissible values (Gerasimenko *et al.* 2002). In its simplest form we can rewrite equations (7) as constrains on precision of elements of vector $\mathbf{e}$ in the following form: $\sigma^2_{\mathbf{e}_i} \le \sigma^2_{\mathbf{e}_i,\max}$ for $i = 1, 2, ..., \mathrm{k}$, where $\sigma_{\mathbf{e}_i}$ are the standard deviations of the least-squares estimates of k deformation parameters and $\sigma_{\mathbf{e}_i,\max}$ are some boundary values of precision for these parameters set by the user.

Finally we formulate the optimal design problem in the following form:

$$\begin{aligned} &\Phi = \sum_{i=1}^{3n(n-1)/2} c_i p_i \to \min; \\ &0 \le p_i \le p_{i,\max},\ i = 1, 2, ..., 3n(n\text{-}1)/2; \\ &\varphi_i(\mathbf{Q}_{\mathbf{e}}) \le \varphi_{i,\max}(\mathbf{Q}_{\mathbf{e}}),\ i = 1, 2, ..., r. \end{aligned} \qquad (8)$$

In the following section we show the application of the method described above using two examples of already working monitoring GPS networks using simple deformation models.

3 Numerical examples

We only consider horizontal components of GPS vectors between all possible pairs in the geodynamic geodetic network because accuracy in the vertical direction is poor and we do not expect any significant vertical motions in the considered networks. The components of individual baselines are assumed to have uniform precision. We also assume that possible station locations are restricted by regional topography, equipment safety requirements and other considerations. Thus GPS stations can only be installed at some specific sites determined before optimization.

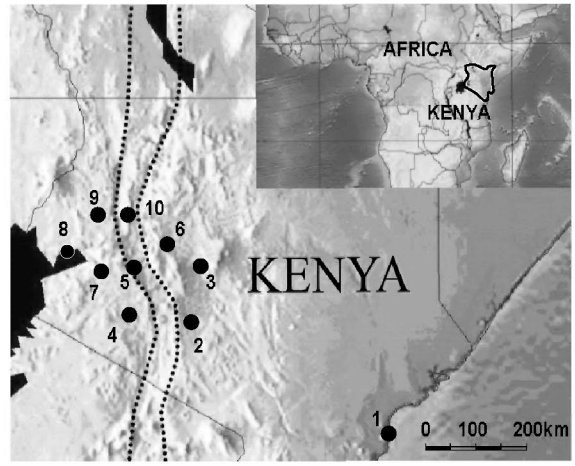


Fig. 1 The Kenya Rift monitoring GPS network. Black circles are possible GPS station locations. Dotted lines are approximate active rift zone margins.

The standard deviation (in meters) of baseline components between network points i and j is given by $\sigma_{ij} = 0.003 + l_{ij} \times 0.05\,\mathrm{ppm}$, where l_{ij} is the baseline length in kilometers. We use this equation in calculating the elements of diagonal weight matrix $\mathbf{Q}^{-1}$ of GPS observations. The reference factor $\sigma_0 = 1$. All coefficients of profitableness $c_i = 1$.

The optimal design problem (8) was implemented in FORTRAN-77 code by the gradient methods with a penalty functions, see for example Gerasimenko (1992).

Example 1. The Kenya Rift monitoring GPS network. The network extends across the Kenya rift (KR) and consists of 10 GPS stations, see Figure 1, which is intended for monitoring the spreading of the KR. The spreading rate of the KR has been reported by Chu and Gordon (1999) within 6 mm/yr in the east-west direction. However values as high as 12 mm/yr has been mentioned elsewhere, Waithaka [personal communication]. In constructing deformation model (4) we assumed that all GPS points are located on two rigid plates (eastern and western) with the eastern plate of the KR being fixed (points 1, 2, 3 and 6 have no movements). Network points 4, 7, 8, 9 are moving in the west direction with relative velocity b (unknown deformation parameter). Points 5 and 10 are situated within the rift basin and we exclude them from this analysis. In this case there are 28 possible baselines that can be measured between network points. We determine the constrain on precision of deformation

parameter b in the form $\sigma_b^2 \le \sigma_{b,\max}^2$. Matrix **B** has dimensions $2n \times 1$ and consists of 1 and 0 elements only. The value of σ_b is determined from equation (5) taking (6) into account. The maximum achievable precision of deformation parameter σ_b with a full set of GPS observations is equal to 4.5 mm/yr.

The solution of the optimization problem (8) under the conditions determined above gives the following results: if the required precision of spreading rate $\sigma_{b,\max} = 12$ mm/yr then baseline 2-4 components provide the required precision, measured with initial (determined before optimization) weights. If $\sigma_{b,\max} = 8$ mm/yr then baselines 2-4 and 6-7 components should be measured with initial weights. In addition components of vector 6-9 should be observed but with their initial weights reduced by a factor of approximately 1.7. In case $\sigma_{b,\max} = 6$ mm/yr, components of the baselines 2-4, 3-4, 4-6, 6-7, 6-9 with their initial weights should be included into the optimal network scheme. If $\sigma_{b,\max} = 5$ mm/yr, then baselines 2-4, 2-7, 3-4, 3-7, 3-9, 4-6, 6-7, 6-8 and 6-9 components should be measured with the weights determined before optimization.

Besides eliminating useless observations with weights close to zero some GPS stations can also be excluded from the optimal network scheme. For example 6 stations can be eliminated if $\sigma_{b,\max} = 12$ mm/yr but only 1 station can be excluded if $\sigma_{b,\max} = 5$ mm/yr. Our results also show that shorter baselines and their east-west components are more important for achieving the required precision of the given spreading rate. Generally this is in agreement with results from other works (Even-Tzur 2002, Stopar 2001). However if we take into account more realistic values of coefficients of profitableness of GPS observations (not equal to unit) it is possible to obtain significantly different optimal projects of the same network.

Assuming that we want to keep baseline 2-9 components through all optimal projects because this baseline has been observed continuously for a long duration or for some other reasons, we can do this by artificially constraining the gradients of this baseline's components to zero during optimization process. In this case, if $\sigma_{b,\max} = 12$ mm/yr we should observe vector 2-9 components with initial

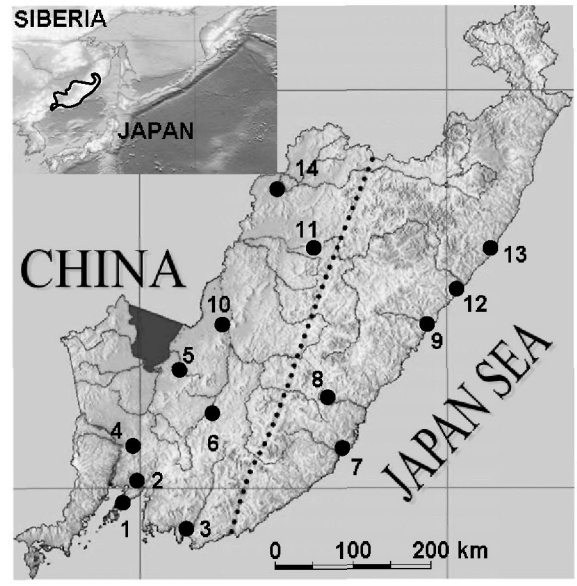


Fig. 2 The Primorsky monitoring GPS network. Black circles are possible GPS station locations. Dotted line is the Central Sikhote-Alinsky Fault trace.

weights and baseline 2-4 components with their initial weights reduced by factor 1.4. If $\sigma_{b,\max} = 8$ mm/yr components of vectors 2-9, 2-4, 6-7 should be observed with their initial weights and components of 6-9 measured with its initial weights reduced by factor 5 respectively. On the other hand if $\sigma_{b,\max} = 6$ mm/yr components of baselines 2-9, 2-4, 4-6, 6-7 and 6-9 should be included in the optimal network scheme with initial weights while the weights of vector 3-4 components are reduced by factor 1.4. For the case where $\sigma_{b,\max} = 5$ mm/yr the same vector components as in the previous case ($\sigma_{b,\max} = 6$ mm/yr) should be observed including additional components from baselines 2-7, 3-7 and 6-8 with initial weights to achieve the required precision of the KR spreading rate.

Example 2. The Primorsky monitoring GPS network. The main goal of this network is monitoring of crustal movements related with the recent geodynamic activity of the Central Sikhote-Alinsky Fault (CSAF). This is a left-lateral strike-slip fault. It extends from the south-west to the north-east along azimuth 25°. There is no recent information about the geodynamic activity of the CSAF. However Ivanov (1972) estimated the probable value of slip-rate from 5 to 7 mm/yr based on prior geological investigations. The short period of our GPS ob-

servations does not allow us to produce more reliable estimate of these movements.

The GPS stations can be only installed at 14 fixed sites, see Figure 2. In total 91 GPS vector can be measured between all network points.

We assumed that the rigid body displacement model (Chrzanowski *et al.* 1983) can be adopted in describing the crustal movements related with the CSAF. In this case the network points are located on two rigid plates (western and eastern) with the western side of the CSAF considered fixed (stations 1 to 6, 10, 11 and 14). Stations 7, 8, 9, 12 and 13 are located on the object side. The slip-rate along the fault trace is the unknown parameter. The N-S and E-W displacements of the object side points are equal to 0.89 and 0.45 of total slip-rate respectively because the azimuth of the fault trace is equal to 25°. Thus, matrix $\mathbf{B}$ has dimensions $2n \times 1$ and consists of 0.00, 0.89 and 0.45 elements only. All other assumptions are as in **example 1**. The maximum achievable precision of slip-rate along the fault trace with a full set of GPS observations is equal to 3.6 mm/yr.

Let's assume that we want to keep components of baselines 1-8, 3-8, 4-8 and 8-11 through all optimal projects of this network. Solution of the non-linear optimization problem gives the following results: with $\sigma_{slip-rate,\max} = 7$ mm/yr we should measure components of the four already observed baselines in addition to components of the GPS vectors 6-8, 8-10, 9-11 with initial weights. For the case $\sigma_{slip-rate,\max} = 5$ mm/yr, baseline 5-7, 5-8, 6-7, 6-8, 7-10, 8-10, 9-11, 11-12, 11-13 components with initial weights and components of baseline 3-7 with weights reduced by factor 1.7 should be added to the set of four vectors components already observed. In case $\sigma_{slip-rate,\max} = 4$ mm/yr the number of additional baseline vector components increases significantly. Baselines 1-7, 2-7, 2-8, 3-7, 4-7, 5-7, 5-8, 6-7, 6-8, 6-9, 7-10, 7-11, 8-10, 8-14, 9-10, 9-11, 9-14, 10-12, 11-12, 11-13, 12-14, 13-14 components should be observed with their initial weights.

Similar to the Kenya GPS network, the shortest baselines are more informative in determining the slip-rate of the CSAF with the required precision. The number of eliminated observations and stations in the optimized network depends on the required precision of deformation parameter. However, with two simple examples we have shown that it is possible to significantly reduce the network campaign cost even if the required precision is close to the maximum achievable precision with the full set of GPS measurements.

4 Conclusions

In the present study the geodetic network design problem under cost criterion was formulated as a non-linear optimization task and solved using well-known mathematical methods. We have also demonstrated its application using two local monitoring GPS networks. Following the proposed methodology, we can eliminate useless observations with weights close to zero from the optimal network project (SOD). In some cases one or more GPS stations can also be excluded from the optimal GPS network scheme (FOD). By artificially setting the gradients of corresponding observations equal to zero, we can keep those measurements throughout all the optimal projects of the network, and only add those observations necessary for achieving the required precision of deformation parameters for a chosen deformation model (TOD). After the optimization we can still include in the final network scheme some observations previously deleted during the optimization say, for redundancy.

In any optimization problem the construction of the variance-covariance matrix of GPS observations used for accuracy estimation of the deformation parameters is always complicated. Therefore some realistic approaches should be used in its construction see, for example Ananga *et al.* (1994). However it may be problematic in the case of strongly physically correlated GPS measurements (Gerasimenko *et al.* 2002; Strang and Borre 1997).

Another issue is determining the correct coefficients of profitableness in the target function. Using different approaches in calculating such coefficients it is possible to obtain different optimal projects of the same geodetic network. Despite these shortcomings in the proposed methodology we think that it can be successfully applied as a useful tool in designing local geodynamic GPS networks.

Acknowledgments. This work was supported by Japan-Russia Youth Exchange Center (Obuchi Fund) and also partially supported by Russian foundation for basic research (grant number: 03-05-64275-a). The authors would also like to thank the anonymous reviewers of this paper for their constructive and helpful remarks.

References

Afonin KF (1975) On Optimal Design of Geodetic Network (in Russian). Proc NIIGAiK, Novosibirsk 37: 141-152

Ananga N, Coleman R, Rizos C (1994) Variance-covariance estimation of GPS networks. Bull Geodesique 68: 77-87

Chrzanowski A, Chen YQ and Secord JM (1983) On the strain analysis of tectonic movements using fault crossing geodetic surveys. Tectonoph 97: 297-315

Chu D and Gordon RG (1999) Evidence for motion between Nubia and Somalia along the Southwest Indian ridge. Nature 368: 64-67

Even-Tzur G (2002) GPS vector configuration design for monitoring deformation networks. J Geod 76: 455-461

Even-Tzur G and Papo HB (1996) Optimization of GPS networks by linear programming. Surv Rev 33: 537-545

Gerasimenko MD (1992) Optimal Design and Adjustment of Geodetic Networks (in Russian). Nauka Press, Moskow

Gerasimenko MD, Kasahara M and Shestakov NV (2002) Optimal Design of Scheme of Deformation Geodetic Networks (in Russian). Geodesy and Air-proto-surveying, Moskow 4: 21-35

Gerasimenko MD, Shestakov NV, Kato T (2000) On optimal geodetic network design for fault-mechanics studies. Earth Planet Space 52: 985-987

Grafarend EW (1974) Optimization of Geodetic networks. Boll Geod Sci Aff 33 (4): 351-406

Ivanov AV (1972) The Central Sikhote-Alinsky Fault (in Russian). Far Eastern Pub Press, Vladivostok

Koch KR (1985) Optimization of the configuration of a network by introducing small position changes. In: Grafarend, Sanso (eds) Optimization and design of Geodetic Networks. Springer, Berlin Heidelberg New York, pp 56-73

Kuang S (1996) Geodetic Network Analysis and Optimal Design: Concepts and Applications. Ann arbor Press, Inc, Chelsea, Michigan

Schmitt G (1982) Optimization of geodetic networks. Rev Geophys Space Phys 20 (4): 877-884

Provorov KL and Afonin KF (1976) Some aspects of optimal design of geodetic networks (in Russian). Geodesy and Cartography 1: 16-20

Schmitt G (1985) Review of Network Designs: Criteria, Risk Functions, Design Ordering. In: Grafarend, Sanso (eds) Optimization and design of Geodetic Networks. Springer, Berlin Heidelberg New York, pp 6-10

Stopar B (2001) Second order design of horizontal GPS net. Surv Rev 36: 44-53

Strang G, Borre K (1997) Linear algebra, geodesy, and GPS. Wellesley-Cambridge Press, Massachusetts

Deformation of the South American Crust Estimated from Finite Element and Collocation Methods

Hermann Drewes
Deutsches Geodätisches Forschungsinstitut (DGFI), Marstallplatz 8, 80539 München, Germany
e-mail: drewes@dgfi.badw.de

Oliver Heidbach
Geophysical Institute, Karlsruhe University, Hertzstr. 16, 76187 Karlsruhe, Germany
e-mail: oliver.heidbach@gpi.uni-karlsruhe.de

Abstract. The present-day surface velocity field of the South American continent reflects the recent geodynamic features. It varies from quasi rigid body motion in the eastern part to mountain range deformation in the western Andes. Space geodetic observations provide velocities at discrete points only. To represent the continuous velocity field, an adequate deformation model has to be developed. Two different models are applied, the least squares collocation approach (LSC) and the finite element method (FEM). The input data are given by 329 velocities derived from continuously observing GPS stations and several GPS geodynamics projects. The different data sets are transformed to a common kinematic datum by deriving the rotation vector of the South American plate from station motions of the IGS Regional Network (RNAAC-SIR) in the rigid eastern part and reducing these plate motions from all the data sets. The resulting residual motions define the boundary conditions in the FEM and the input signals in the LSC. For the FEM a network of approximately 75000 linear elements is generated. The model rheology is a homogeneous elastic material (Young's modulus 70 GPa and Poisson ratio 0,3). For the LSC empirical covariance functions are derived from the observed velocity vectors. The comparison of both methods shows an agreement in the mm/a level. The result is a continuous surface velocity model for the South American continent which may be used for interpolation of point motions in geodetic networks and reference frames. A detailed geodynamic interpretation has not yet been done due to the simplicity of the model (homogeneous material, rough fault structures, no sophisticated dynamics).

Keywords. Recent crustal movements, South America, SIRGAS, GPS velocity field, collocation, finite element method

1 Introduction

The deformation process of the South American continent is mainly controlled through the push of the Mid Atlantic Ridge from the east, the subduction of the Nazca plate along the Pacific coast in the west and the subduction of the Caribbean plate in the north (fig. 1).

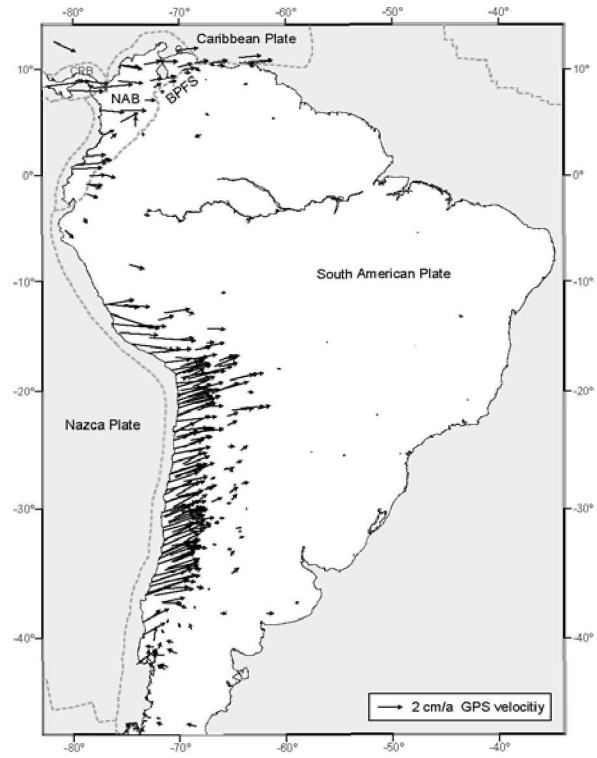


Fig.1: Velocities relative to the stable part of South America derived from different GPS projects (see text). Plate boundaries are taken from (Bird, 2003). NAB = North Andes Block, BPFS = Boconó-El Pilar fault system CRB = Costa Rica Block.

The North Andes Block in the north-western part of the continent (Audemard and Audemard, 2002; Kellogg et al., 1995) is bounded to the south by the dextral shear along the Boconó-El Pilar fault system. As a result the present-day velocity field, as observed by space geodetic techniques, shows a strong variation from east to west (fig. 1).

Dislocation models based on the boundary element method show that the interface between the subducting Nazca plate and the overriding South America plate is locked (Bevis et al., 2001; Khazaradze and Klotz, 2003). Therefore, the movement of the GPS stations in magnitude and direction is mainly due to the relative movement of the Nazca plate with respect to the stable part of the South America plate. This relative movement of the Nazca plate has a present velocity of 65 mm/a with an azimuth of N77°E at 30°S (Angermann et al., 1999). The dislocation modelling results (Bevis et al., 2001; Khazaradze and Klotz, 2003) show that most of the observed inter-seismic velocity field can be explained by that transmission of deformation. The velocity is decreasing perpendicularly to the trench and comes close to zero east of the Sub-Andes where the stable part of the South America plate begins.

In the northern part of the South American plate the situation is more complex. Here the north-west striking and south-east dipping subduction of the Caribbean plate and the collision of the Costa-Rica block influence the deformation of the subduction process described above (Audemard and Audemard, 2002; Kellogg et al., 1995). As a result the North Andes block has been developed. This block moves along a transpressive dextral strike-slip fault system in eastern direction with 10-20 mm/a (Kaniuth et al., 1999).

The South American Geocentric Reference System, SIRGAS (Hoyer et al., 1997; Drewes et al., 2000) defines the station coordinates at a fixed epoch (1995.4 or 2000.4, respectively). To reduce the coordinates determined at another epoch to the reference epoch, the station velocities are needed (Drewes et al., 1997). These may be derived for the SIRGAS stations from repeated observations (1995 and 2000). For new stations, e.g., in the national reference frames, a continuous velocity field is needed to interpolate the required velocities.

2 Used Data Sets and Reductions

Point velocities in South America are available from the following data sets of GPS observations: The IGS Regional Network Associate Analysis Centre for SIRGAS, RNAAC-SIR (Seemüller et al. 2002), the SIRGAS campaigns 1995 and 2000 (Luz et al., 2000, Drewes et al. 2004), the Central and South America geodynamics project, CASA (Kaniuth et al., 1999), the South America - Nazca Plate Motion Project, SNAPP (Norabuena et al., 1998), the Central Andes Project, CAP (Brooks et al.. 2003, Kendrick et al., 1999; Kendrick et al., 2001, Lamb 2000) and the South America Geodynamic Activities, SAGA (Khazaradze and Klotz, 2003; Klotz et al., 2001).

In a first step, the rotation pole of the rigid South American plate was estimated from the velocities of the RNAAC-SIR stations in the stable western continent. The RNAAC-SIR velocities, originally referring to the ITRF 2000 reference frame, were reduced by the plate velocities resulting thus in velocities relative to the rigid South American plate.

In the second step, the velocities of the non-continuously observing SIRGAS stations were derived from the coordinates of the SIRGAS 1995 (transformed from the original ITRF94 to ITRF 2000) and the SIRGAS 2000 (given in ITRF 2000) campaigns. These velocities were then transformed to the RNAAC datum by identical stations.

In the following steps, the velocities of all the geodynamics projects were transformed from their individual kinematic datums (in general relative to an arbitrarily defined stable South America) to the RNAAC datum by identical stations.

The accuracy of observed velocities is in general in the order of some millimeters per year (mm/a). Observations at stations affected by earthquakes were neglected, i.e., only data which reflect the inter-seismic velocity field were taken into account.

3 Modelling of the Velocity Field

To model the velocity field we applied two different methods: the least squares collocation approach and the finite element method. The first one is a mathematical vector prediction algorithm without any geophysical models. It interpolates the velocity field from the discrete observations. The second method uses a simple geophysical model of plate tectonics with an deformable elastic material.

3.1 Finite Element Model

The **V**elocity **Mo**del of **S**outh America (VEMOS) from the finite element approach consists of 75000 linear 3D-elements on a sphere with a constant radius of 6371 km. The equation of equilibrium is solved for simple homogeneous linear elastic rheology. The Young module was set to 70 GPa and the Poisson number is 0,3. In the north, the Boconó-El-Pilar fault system which marks the southern border of the North Andes block is implemented as a contact surface where relative sliding is allowed.

The coefficient of friction is set to 0,1. At the northern (20°N) and southern (55°S) boundaries sliding in west-east direction is allowed, the western boundary (-90°W) is free of conditions and the eastern boundary (-34°W) is fixed. The bottom of the model at 10 km depth is fixed in the vertical direction; the surface of the model is free (fig.2). Further boundary conditions are the horizontal components of the given 329 GPS velocities (fig.1).

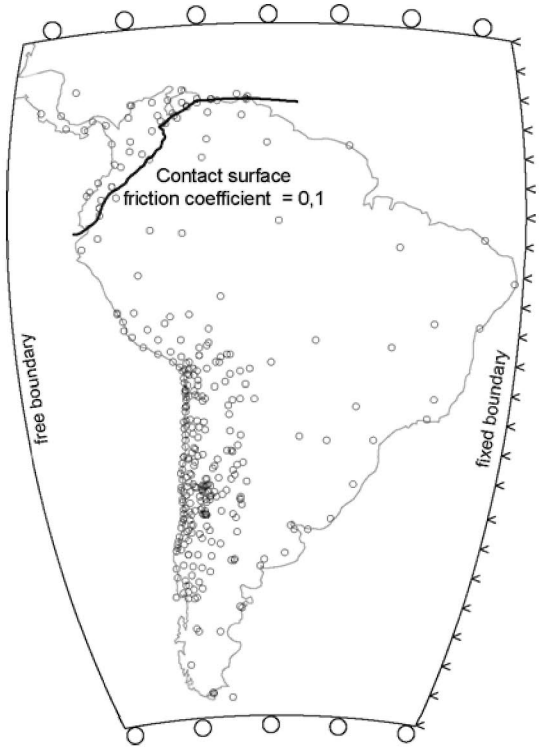


Fig.2: Boundary conditions for the finite element model: The northern and southern boundary condition allows east-west sliding and the eastern boundary is fixed. The bottom of the model at 10 km depth is fixed in vertical direction; the surface of the model is free. The Boconó-El-Pilar fault system (grey line) is modelled as a contact surface where relative sliding is allowed. The circles are the locations of the GPS observation points.

3.2 Least Squares Collocation

The least squares collocation approach uses the well-known relation between correlated signals:

$$\mathbf{v}_{pred} = \mathbf{c}_{in}^{T}\, \mathbf{C}_{ij}^{-1}\, \mathbf{v}_{obs}$$

$\mathbf{v}_{obs}$ are the observed and $\mathbf{v}_{pred}$ the predicted point velocity vectors, $\mathbf{C}_{ij}$ is the auto-covariance matrix between the observed and $\mathbf{c}_{in}$ the cross-covariance matrix between the observed and the predicted velocity vectors.

The elements of the covariance matrices are taken from empirically determined isotropic covariance functions, each one for the north and east velocity components and one for the cross correlation between north and east components. The empirical correlations are computed in distance (d) classes between the points and approximated by a simple exponential function: $a \cdot \exp(-b \cdot d)$. To get a sufficiently dense input velocity field, a wide-spaced grid is interpolated linearly in areas with sparse observations, in particular in the central part of the continent (Amazon region).

The prediction error of velocities is typically in the order of the observation errors, i.e., some mm/a.

4 Results

4.1 Finite Element Model

The results of the finite element model (fig. 3) can be distinguished in four first order patterns:
(1) The high gradient of the velocity field across the dextral Boconó-El-Pilar fault system in the north-west of the continent. This gradient decreases from east to south-west due to increasing transpressional

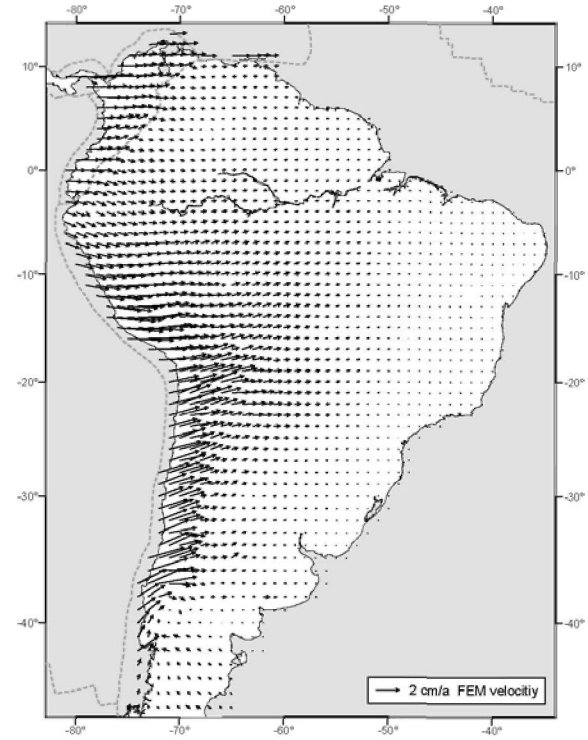


Fig.3: Horizontal velocity field from the finite element model with a Young module of 70 GPa and a Poisson number of 0,3. The coefficient of friction at the Boconó-El-Pilar fault system is set to 0,1. Boundary conditions at the model border and the GPS data are presented in the previous figures.

(2) The area between the equatorial region and 25° southern latitude where the velocity field decreases smoothly east of the Sub-Andes. The decrease is nearly linear from west to east and becomes zero at the Atlantic coast
(3) The area between 25°S and 38°S where the velocity field decreases east of the eastern cordillera and trends then smoothly to zero at the Atlantic coast.
(4) The area south of 37°S where the few observed velocity vectors do not show a clear picture and therefore the velocity field is not quite significant.

4.2 Least Squares Collocation

The result of the collocation approach (fig. 4) shows naturally in principle the same four mentioned patterns, but it deviates slightly in pattern (2). Here the decrease is stronger than in the finite element model. This may be due to the steep covariance function in this region which is characteristic for short correlation lengths between the velocities. For this reason the large velocities in the Andean deformation zone are not transferred to the central part of the continent. That means, that we have a relatively sharp separation between deformation zone and rigid plate.

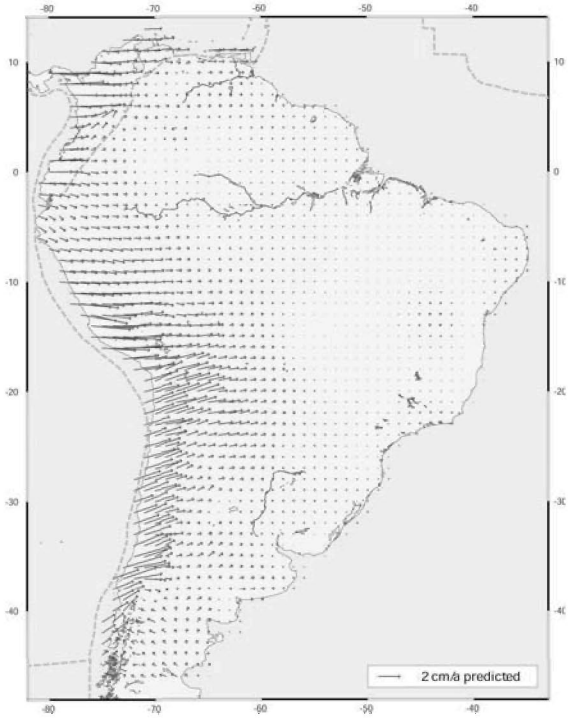


Fig.4: Horizontal velocity field from the least squares collocation (vector prediction) approach with empirically determined covariance functions.

5 Comparison

The numerical comparison of the two approaches shows a general good agreement. The numbers are:

Average difference West-East:	– 1.0 mm/a
Average difference South-North	– 0.1 mm/a
r.m.s. difference West-East:	± 1.7 mm/a
r.m.s. difference South-North:	± 0.9 mm/a.

The main systematic deviations between both models are generated by the smooth (FEM) or steep (LSC) transition from the Andean deformation zone to the rigid region in the central continent. The differences come up to a maximum of 7 mm/a in West-East and 5 mm/a in South-North direction. This phenomenon has further to be investigated, in particular with respect to the elasticity parameters and the covariance functions.

6 Discussion

The result of this research project is a continuous surface velocity model for the South American continent which may be used for interpolation of point motions in geodetic networks and reference frames. The long-term problem is that the used data sets only reflect the inter-seismic signals, i.e., in regions where major earthquakes occur the reduction of coordinates cannot be done with linear velocities since the co- and post-seismic signal changes the velocity field in a wide area (Angermann et al., 2003; Kaniuth et al., 2002; Pollitz et al., 2000). Since all the GPS stations in the western part of the continent close to the Nazca subduction will sooner or later be affected by earthquakes, the velocity models are only a first approximation.

Also the effects of further major strike slip and thrust faults in the Andes were neglected and are probably not accurately reflected in the GPS data. For future demands, either from scientific curiosity or from geodesy, on higher resolutions in time and space the finite element approach seems to be the better option than the collocation approach. The simple kinematic model with homogeneous elastic material, only rough fault structures and no sophisticated dynamics, however, has to be improved and to be extended to a complete three-dimensional dynamic viscous-elastic-plastic model with detailed tectonic structures (e.g. Heidbach and Drewes, 2003). The more GPS data become available the more we will learn on relative movements along active faults or rigid blocks which move along major fault systems. In the present study we did not accomplish a detailed geodynamic interpretation due to the simplicity of the models.

7 Conclusion

We conclude three statements:
(1) More observations in space and time are needed to isolate the linear from the non-linear part of the observed signals. There are observation gaps, e.g., between 0° and 10°S and in southern Chile,
(2) For more accurate application of the velocity field the influence of the viscous relaxation processes in return of the earthquakes must be investigated, and
(3) More neotectonic structures and a more sophisticated rheology to account for the relaxation processes has to be applied in a more detailed model which incorporates, e.g., thrust faults of the Sub-Andes (Norabuena et al., 1998) and inhomogeneous and layered rheolgy (Wdowinski and Bock, 1994).

8 References

Angermann D., Klotz J., Reigber C. (1999) Space-geodetic estimation of the Nazca-South America Euler vector. *Earth Planet Sc Lett.* **171**, 329-334.

Angermann D., Krügel M., Meisel B., Müller H., Tesmer V. (2003) Time series of station positions and datum parameters. *Geotechnologien Science Report* **No. 3,** 17-21.

Audemard F.E., Audemard F.A. (2002) Structure of the Mérida Andes, Venezuela: relations with the South America-Caribbean geodynamic interaction. *Tectonophysics* **345**, 299-327.

Bevis M., Kendrick E., Smalley J.R., Brooks B., Allmendinger R., Isacks B. (2001) On the strength of interplate coupling and the rate of back arc convergence in the central Andes: An analysis of the interseismic velocity field. *Geochem., Geophys., Geosyst.* **2**, 10.1029/2001GC000198.

Bird P. (2003) An updated digital model for plate boundaries. *Geochem., Geophys., Geosyst.* **4** (3), 1027, doi: 10.1029/2001GC000252.

Brooks, B.A., Bevis, M., Smalley, J.R., Kendrik, E., Manceda, R., Lauría, E., Maturana, R., Araujo, M. (2003) Do the Andes behave like a microplate? *G^3 Geochem., Geophys., Geosyst.* **4**, (10), 1085, doi: 10.1029/2003GC000505.

Drewes H. (1997) Time evolution of the SIRGAS reference frame. *IAG Symposia* **118,** 174-179, Springer, Berlin.

Drewes H., Kaniuth K., Seemüller W., Stuber K., Tremel H., Hernandez N., Hoyer M., Wildermann E., Fortes L.P., Pereira K.D. (2000) Monitoring the Continental Reference Frame in South America. *IAG Symposia* **124**, 134-137. Springer, Berlin.

Drewes, H., Kaniuth, K., Völksen, C., Costa, S.M., Fortes, L.P. (2004) Results of the SIRGAS Campaign 2000 and coordinates variations with respect to the 1995 South American Geocentric Reference Frame. *IAG Symposia* (this volume) Springer, Berlin.

Heidbach, O., Drewes, H. (2003) 3-D Finite Element model of major tectonic processes in the Eastern Mediterranean. In: Nieuwland, D.A. (ed.) New insigths in structural interpretation and modelling, *Geological Society, Spec. Pub.* **212**, 261-274, London.

Hoyer, M., Arciniegas, S., Pereira, K., Fagard, H., Maturana, R., Torchetti, R., Drewes, H., Kumar, M., Seeber, G. (1997) The definition and realization of the reference system in the SIRGAS project. *IAG Symposia* **118,** 168-173, Springer, Berlin.

Kaniuth K., Drewes H., Stuber K., Tremel H., Hernández N., Hoyer M., Wildermann E., Kahle H. G., Geiger A., Straub C. (1999) Position changes due to recent crustal deformations along the Caribbean - South American plate boundary derived from the CASA GPS Project. *XXII IUGG General Assembly*, Birmingham, UK.

Kaniuth K., Müller H., Seemüller W. (2002) Displacement of the space geodetic observatory Arequipa due to recent earthquakes. *Zeitschr. für Verm.* **127**(4), 238-243.

Kellogg J.N., Vega V., Stallings T.C., Aiken C. L.V. (1995) Tectonic development of Panama, Costa Rica, and the Colombian Andes: Constraints from Global Positioning System geodetic studies and gravity. In: Mann, P. (ed.) Geologic and Tectonic Development of the Caribbean Plate Boundary in Southern Central America, *Geol. Soc. Am. Special Paper* **295,** 75-90.

Kendrick E., Bevis M., Smalley J.R., Brooks B. (2001) An integrated crustal velocity field for the central Andes. *Geochem. Geophys. Geosyst.* **2**, 10.129/2001GC000191.

Kendrick E.C., Bevis M., Cifuentes O., Galban F. (1999) Current rates from convergence across the Central Andes: Estimates from continuous GPS observations. *Geophys. Res. Lett.* **26**(5), 541-544.

Khazaradze G., Klotz J. (2003) Short- and long-term effects of GPS measured crustal deformation rates along the south central Andes. *J. Geophys. Res.* **108** (B6), 2289, doi: 10.1029/2002JB001879.

Klotz J., Khazaradze G., Angermann D., Reigber C., Perdomo R., Cifuentes, O. (2001) Earthquake cycle dominates contemporary crustal deformation in Central and Southern Andes. pp. 437-446.

Lamb S. (2000) Active deformation in the Bolivian Andes, South America. *J. Geophys. Res.* **105**, 25627-25653.

Luz, R.T., Fortes, L.P., Hoyer, M., Drewes, H. (2002) The vertical reference frame for the Americas – The SIRGAS 2000 GPS campaign. *IAG Symposia* **124**, 302-305. Springer, Berlin.

Meijer P.T. and Wortel M.J.R. (1997) Present-day dynamics of the Agean region: A model analysis ot the horizontal pattern of stress and deformation. *Tectonics* **16**, 879-895.

Norabuena,E., L. Leffler-Griffin; A. Mao; T. Dixon; S. Stein; I.S. Sacks; L. Ocola; M. Ellis (1998): Space geodetic observations of Nazca-South America convergence across the Central Andes. *Science* (279) 358-362.

Pollitz F. F., Peltzer G. and Bürgmann R. (2000) Mobility of continental mantle: Evidence from postseismic geodetic observations following the 1992 Landers earthquake. *J. Geophys. Res.* **105**, 8035-8054.

Seemüller, W., Kaniuth, K., Drewes, H. (2002) Velocity estimates of IGS RNAAC SIRGAS stations. *IAG Symposia* **124**, 7-10. Springer, Berlin.

Wdowinski S. and Bock Y. (1994) The evolution of deformation and topography of high elevated plateaus: 2. Application to the central Andes. *J. Geophys. Res.* **99**, 7121-7130.

Present-day Movement and Tectonic Strain Rate Field of Active Crustal Blocks in China Continent

Caijun XU , Hua WANG (School of Geodesy and Geomatics, The Key Laboratory of Geospace Environment and Geodesy, Ministry of Education, Wuhan University, Wuhan, 430079, P. R. China, cjxu@sgg.wtusm.edu.cn)

Jusheng XU(Institute of Seismology, China Seismological Bureau, Wuhan, 430071, P.R.China)

Abstract: China continent, constructed by many blocks, is a dynamical system with complicated geological structure, topography and physiognomy. Since Cenozoic, it takes on different tectonic deformation in different blocks and the relative movement between neighboring blocks does always exist, which may be the key inducement of the strong seismic activity at the inner of China continent (Zhang Peizhen , 1999). In terms of tectonic movement feature, the continent can be divided into several different scalar crustal and lithosphere blocks, which include 6 active block regions that are Qinghai-Tibet, Xingjiang, Northeast, North China, South China and South China Sea block region, two active plate boundary tectonic zones and 29 active blocks that are Lhasa, Tian Shan etc. (Deng Qidong, Zhang Peizhen *et al.*, 2003). In this paper, based on the bi-cubic Bessel spline function method, we inverted the present-day movement field of active blocks in China continent by combining with GPS measurements, seismic moment tensors of earthquakes and fault slip rates in China and its neighboring areas, considering about the geological and geophysical parameters at the same time. Furthermore, we inverted the strain rate field, strain energy density rate field ,then we calculated the shearing strain rate field, area strain rate field and analyzed the relationship between the strong seismic activity and strain rate fields. The results reveal that in the middle and west of China continent, tectonic activity is stronger, particularly along the boundary zone of Himalayan block, in ChuanDian area of Qinghai-Tibet block and Tian Shan area of Xingjiang block, shearing strain rate is larger, the shearing strain rate fields are in correspondence with the tendency of present-day tectonic movement; from the distribution of principal axes of tectonic stress field, we can infer that present-day movement of active blocks in China continent is mainly caused by posthumous movement. The continent of China deformation has not only the strike-slip faulting and block's transverse slip feature but also the crustal shortening and thickening feature, it doesn't support the continental eastward escape theory of high-rate of slip.

Key Words: Tectonic strain & stress field, Active fault block, China continent

1. Tectonic Background

China continent situates in the southeast of Eurasia plate, which is surrounded by Indian plate, Pacific plate and Philippine plate. Many strong tectonic activities have happened since late Quaternary. Himalayan and Taiwan tectonic zone forms the active tectonic boundary and the other Blocks located at the inner of the continent. Tectonic boundary is the most important active tectonic zone since most of present-day orogenics, strong earthquakes and volcanic zones are formed in these areas. Even at the inner of the plate, it is not really rigid, especially that relative movement always exists between neighbor blocks and their extents are different, which formed the active blocks and tectonic zones. Based on the accomplishment of active tectonic quantitative research obtained by Chinese scientists in recent 20 years, China continent is subdivided into six block regions and two active plate boundary tectonic zones .Six block regions could be ulteriorly subdivided into 29 active blocks, they are Lhasa, Qiangtang, Tianshan, ChuanDian , Ordos, Sulu and Songliao block et al.(Deng QD et al.,2003)

2. GPS Observations and Its Velocity Field

In this paper, GPS velocity field is mainly determined by GPS2000 network, as is so called Three Network. The network includes the national GPS Order-A and Order-B network, the national

GPS Order-I and Order-II network, and the Curstal Movement Observation Network of China (CMONC).The national GPS Order-A and Order-B networks have 818 GPS stations, which were distributed by National Bureau of Surveying and Mapping of China from 1991 to 1995. The national GPS Order-I and Order-II distributed by the Bureau of Surveying and Mapping under the PLA headquarters of General Staff from 1991 to 1997 have 534 GPS stations, which were well distributed in all regions of China (except Taiwan). CMONC includes reference network, basic network and regional network, which are distributed by China Seismological Bureau, Bureau of Surveying and Mapping under the PLA headquarters of General Staff, National Bureau of Surveying and Mapping of China and Chinese Academy of Sciences. 1081 GPS stations have been built up in CMONC. The reference network is constituted by 25 continuous IGS track station. Fifty-six GPS stations in the basic network are observed once a year. One thousand GPS stations in the regional networks have been observed twice in 1999 and 2001 respectively. The network adjustment has been accomplished until early 2003. After the adjustment, the relative position precision is up to 10^{-7} -10^{-8}.

In this paper, we mainly use the data from the national high precision GPS order-A and order-B network, Qinghai-Tibet crustal deformation monitoring network, GPS network in the Chinese Southeast Coast, Xinjiang network ,North China network and CMONC et al. The multi-epoch regional GPS data are merged into ITRF2000 at a unity epoch (Wang Qi etal.,2001, Wang M.et al.,2003,Li Yanxing et al.,2003, the detailed description of the processing procedure is given by Wang Qi et al., 2001 and Liu Jingnan etal.,2001). After the unity adjustment, velocities of 1173 stations in ITRF2000 are determined (cf. Fig. 1).

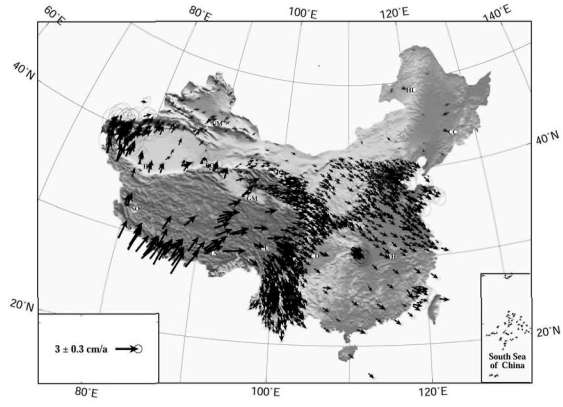


Fig.1 Observation crustal movement velocity field in China continent

3. Joint Inversion Model of Tectonic Strain & Stress Fields

Based on the theory that Asian movement is consistent with fluid slow flow by the effect of gravity, England and Molnar thought that the deformation could be described as the movement of the continuous viscous media affected by gravity(England and Molnar,1997a,1997b). These Asian crustal movement and deformation field could be regarded as a continuous field, that the continental lithosphere of Asia is more properly regarded as belonging to the fluid portion of solid earth than to the relatively small fraction of Earth that behaves as rigid plates. The field was ever fitted with bicubic spline function model. As it is a subsection polynomial and has a good continuity between the subsections, the spline function not only succeed to the simplicity and approach feature of the polynomial but also retain the relative independence among the subsections. So it can reflect the local features better than the normal polynomial model(Haines et al., 1998).

The objective function that is minimized in the inversion is

$$\sum_{i=1}^{N}(\dot{\varepsilon}_{ij}^{fit}-\dot{\varepsilon}_{ij}^{obs})(V)_{ij,pq}^{-1}(\dot{\varepsilon}_{pq}^{fit}-\dot{\varepsilon}_{pq}^{obs}) + \sum_{i=1}^{M}(u_i^{fit}-u_i^{obs})C_{ij}^{-1}(u_j^{fit}-u_j^{obs})$$

where $\dot{\varepsilon}_{ij}^{fit}$ is the average strain rate fitted by the bi-cubic Bessel spline function, $\dot{\varepsilon}_{ij}^{obs}$ is the average strain rate inferred with seismic moment tensors of earthquakes, u_i^{fit} is the velocity field fitted by the bi-cubic Bessel spline function, u_j^{obs} is the velocity field observed by geodetic means such as GPS, $(V)_{ij,pq}^{-1}$ is the priori variance-covariance matrix of the inferred strain rates and C_{ij}^{-1} is the variance-covariance matrix of the GPS velocity observations (the detailed description of the inversion procedure is given by Haines et al., 1998 and Xu Caijun et al., 2002).

In this paper, we use the priori strain rate variance-covariance matrix $(V)_{ij,pq}^{-1}$ to constrain the style and magnitude of deformation. Just like an anisotropic viscous tensor as a whole, $(V)_{ij,pq}^{-1}$ corresponds to the viscosity of regional materials in the given area and controls the strength and anisotropy of each unit medium. By adjusting the priori matrix, we can get all kinds of boundary

conditions and the deformation reflection of the media under the preliminary conditions. The matrix can be expressed as following(Haines et al., 1998):

$$var(\dot{\varepsilon}_{ij}) = [\sum_{k=1}^{N} (\frac{M_{ij}^{k}}{2\mu VT})^2] + \sigma_0^2 \frac{1}{2}(1+\delta_{ij})$$

$$cov(\dot{\varepsilon}_{ij}, \dot{\varepsilon}_{lm}) = [\sum_{k=1}^{N} (\frac{M_{ij}^{k}}{2\mu VT})(\frac{M_{lm}^{k}}{2\mu VT})]$$

where $\sigma_0^2 \frac{1}{2}(1+\delta_{ij})$ is the contribution caused by the "missing" earthquake, $\sigma_0^2 = \frac{\eta h}{V} = \frac{\eta}{S}$, η is the incompleteness factor, which is a priori factor for the purpose of completing the missing earthquake or GPS data. It is often determined by fault slip rates, S is the area of each cell, the priori weight of GPS velocities can be got by the variance-covariance matrix of their observations.

After obtaining $\dot{\varepsilon}_{ij}^{fit}$,we can calculate the 1st and 2nd shear strain rate($\dot{\gamma}_1$, $\dot{\gamma}_2$),maximal shear strain rate($\dot{\gamma}$), area strain rate ($\dot{\Delta}$)and ration rate($\dot{\omega}$) as follows(Draw A.R. et al.,1989, Hu MC and Lu Fu,1994):

1st shear strain rate : $\dot{\gamma}_1 = \dot{\varepsilon}_{EE} - \dot{\varepsilon}_{NN}$

2nd shear strain rate: $\dot{\gamma}_2 = \dot{\varepsilon}_{EN} + \dot{\varepsilon}_{NE}$

maximal shear strain rate: $\dot{\gamma} = (\dot{\gamma}_1^2 + \dot{\gamma}_2^2)^{1/2}$

area strain rate : $\dot{\Delta} = \dot{\varepsilon}_{NN} + \dot{\varepsilon}_{EE}$

ration rate: $\dot{\omega} = (\dot{\varepsilon}_{EN} - \dot{\varepsilon}_{NE})/2$

where E denotes longitude, N denotes latitude, 1st shear strain rate showes pure shear strain part of extending in E-W direction and compressing in S-N direction of strain, 2nd shear strain rate showes pure shear strain part of extending in NE-SWdirection and compressing in NW-SE direction of strain.

4. Inversion Data Selection

To use the bi-cubic spline method to invert the velocity and strain rate fields, the first step is to subdivide the region into some cells. Then the rotation parameters of each knot point can be calculated. Only when all of these steps having been completed, can we apply the spline method in each cell. So the subdivision of grid plays an important role in the inversion. In this paper, an irregular grid has been chosen to sketch the tectonic zone and plate boundaries. The China continent has been subdivided into 279 cells (cf. Fig. 2). In the light of earthquake mechanism at the source and the distribution of the quaternary main active fault, we try to make each cell have the same geological feature. As for some rigid blocks, such as South China, North China and Northeast block, the cells are sparse. On the contrary, as for some complex areas, the cells are dense. In this grid, the Siberia block and the south India plate are regarded as rigid by giving the cells the same rotation parameters in these two blocks. The Siberia block is chosen as the reference frame and all of the velocities are relative to this reference frame. In this inversion, 911 seismic moment tensor data have been used which mainly come from Holt(Holt et al,1995) and Harvard seismology (http: //www. seismicity. Harvard.edu.com) from 1902 to 2003. The velocity observations of 1173 GPS stations are also used(cf Fig.1). Active fault slip rates come from Hong Hanjing(Hong Hanjing et al.1998) and Ren Jinwei(Ren Jinwei ,2001).

The principal axes of strain rate field, area strain rate field, maximal shear strain rate field , 1st , 2nd shear strain rate field and ration rate field inverted are listed by Fig.3 – Fig.8.

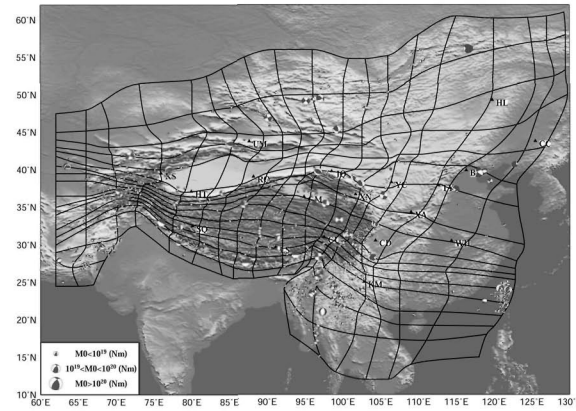


Fig.2 A grid of model

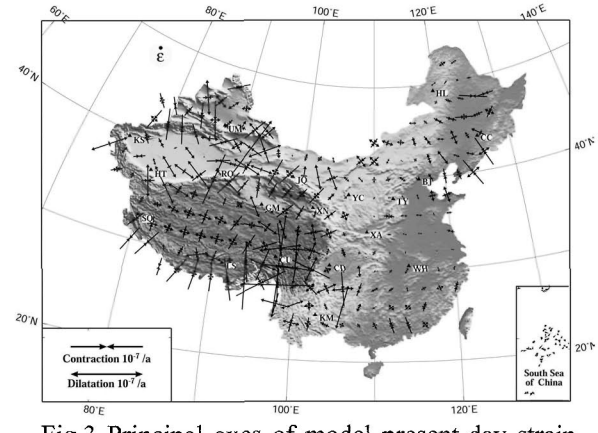


Fig.3 Principal axes of model present-day strain rate field in China continent

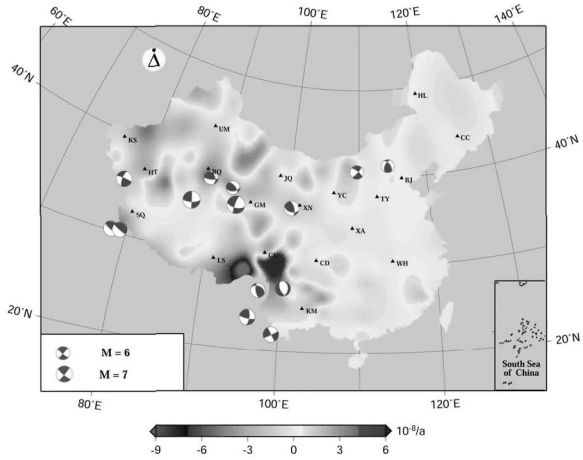


Fig.4 Model present-day area strain rate field in China continent

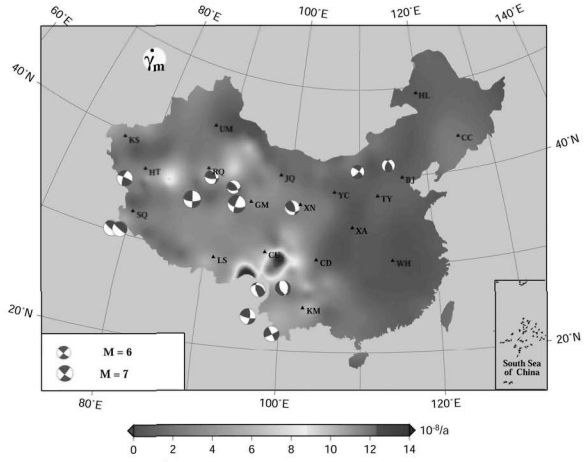


Fig.5 Model present-day maximal shear strain rate field in China continent

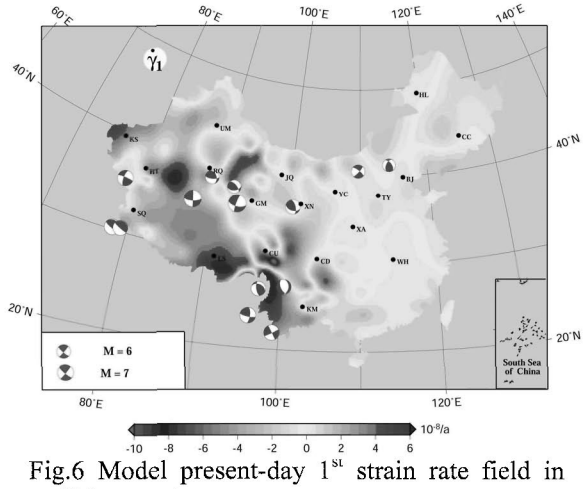


Fig.6 Model present-day 1st strain rate field in China continent

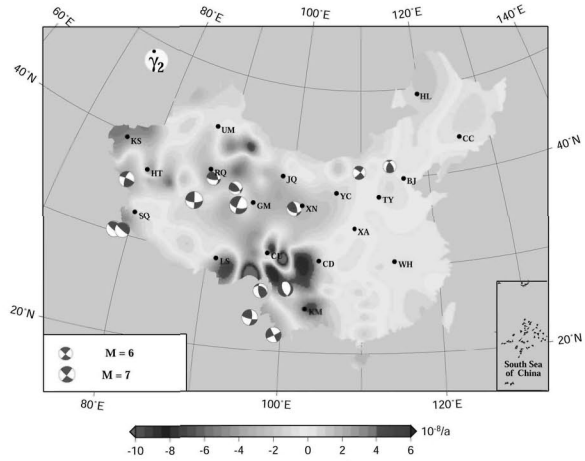


Fig.7 Model present-day 2nd strain rate field in China continent

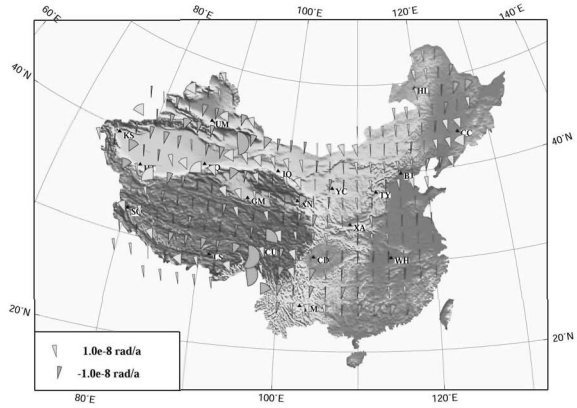


Fig.8 Model present-day rotation rate field in China continent

5. Discussion and Conclusions

From the fig.3 and fig.4, the minimax compressive strain rate is up to -7.1×10^{-8}/yr in Himalayan tectonic zone and -5.5×10^{-8}/yr in Lhasa block. In Qiangtang and Bayankala block, it mainly takes on extension and the maximum value is up to 2.5×10^{-8}/yr. In West Kunlun block, it also mainly takes on extension and the maximum value is up to 7.4×10^{-8}/yr. In Xinjiang block region, it mainly takes on compression and the minimax compressive strain rate is up to -2.4×10^{-8}/yr which happened in Tian Shan block. It mainly takes on extension in Northeast block region. In North China block region, it mainly takes on extension, the maximum value are up to 4.7×10^{-8}/yr, 4.3×10^{-8}/yr, 1.0×10^{-8}/yr and 8.2×10^{-8}/yr in Ordos, Taihang Shan, Hehuai Plain and Jiaoliao block respectively. However, in

North China Plain and Yin-Yan block, it mainly takes on extrusion, the minimax compressive strain rate is up to -7.6×10^{-8}/yr, which happened in North China Plain block. In South China block region, it mainly takes on compressive strain, except in Yangtse River's Middle-lower Reaches where the maximal extension rate is up to 5.9×10^{-8}/yr.

Despite the faultiness of GPS distributions in China continent and the spatial distribution of strain rate field is although a preliminary result, the spatial distribution of strong earthquakes still has some relationship with area strain rate field(cf fig.4). Since 1990s, the strong earthquakes are almost happened on the boundary of expansile and compressive strain rate.

The fig.5 showes that the high shear strain rate focuses on the whole Qinghai-Tibet block and its boundaries and the most part of Xinjiang block region. The highest shear strain rate happened in ChuanDian block, whose value is up to 5.4×10^{-8}. The shear strain rate in the west of China continent is obviously higher than that in the east. In most parts of west, the shear strain rate is higher than 2.0×10^{-8}, however, it is lower than 1.5×10^{-8} in most parts of east.

The fig.6 showes the first shear strain rate is more higher than zero in Lhasa, Himalaya, West-Kunlun, Tianshan, Dianxinan block, they take on extension in E-W dircetion and contraction in S-N direction. It is less than zero in Bayankala , Qiangtang,Altai,North China,ChuanDian and Sulu block, they take on extension in S-N direction and compression in E-W direction

In most of blocks in China continent, the second shear strain rate is negative(cf.fig.7). This result showes that the nearly east-westerward faults show sinistral shear deformation, while the nearly south-northward faults show dextral shear deformation.

From fig.8, we can know that it rotates clockwise in Lhasa Block, Qiangtang, Bayankala block of Qinghai-Tibet block region, these blocks rotate at about 0.4°/m.y. to 5.7°/m.y. While in East Kunlun, Qilian Shan, Daxinanling and Songliao block, it takes on counterclockwise rotation among 0.2°/m.y. to 1.2°/m.y. Ordos block has an anticlockwise rotation rate of about 0.1°/m.y. Taihang Mountain and North China Plain block has an anticlockwise rotation rate of about 0.4°/m.y. to 0.6°/m.y., the rotation situation about North China area is similar to in rotation sense to the result from paleomagnetic studies(Xu Xiwei et al.,1994).

The distribution of principal axes of tectonic strain and stress rate field reveal that different blocks have different deformation in China continent, we can infer that present-day movement of active Blocks in China continent is mainly caused by posthumous movement, which is affected on Pacific, Eurasia, Philippine Sea, India and Australia plates and which is mainly affected by colliding of India Plate with Eurasia Plate. The continent of China deformation has not only the strike-slip faulting and block's transverse slip feature but also the crustal shortening and thickening feature. Both the North China and the South China blocks, east of the Tibetan Plateau, move eastward in significantly slower velocities than that predicted by "extrusion" model. ChuanDian and Dianxinan blocks have smaller clockwise rotation rates than that of Lhasa Block or Qiangtang block. It doesn't support the continental eastward escape theory of high-rate of slip.

Acknowledgment: We deeply express our sincere thanks to our colleagues from China Seismological Bureau, Chinese Academy of Sciences, and National Bureau of Surveying and Mapping, Wuhan University for performing the GPS field measurement supporting this study. Prof. Clark R. Wilson of Geol Science Dept of The University of Texas at Austin and another anonymous reviewer gave useful comments and important suggestions to the authors to improve the manuscript. This work has been supported by National Science Foundation of China (49904001); it has also been supported by Hubei Provincial Outstanding Young Scientists Foundation (2002AC011), and funded By Open Research Fund Program of The Key Laboratory of Geospace Environment and Geodesy, Ministry of Education (02-09-06). Maps were prepared using GMT version3.4 by Wessel and Smith.

References

Deng Qidong, ZHANG Peizhen, RAN Yongkang et al, Basic characteristics of active tectonics of China, Science in China Series D, 46(4): 356-372, 2003.

Drew A.R. et al.:DYNAP:Software for estimating crustal deformation from geodetic data, Tectonophysics, 162, 331-343, 1989

England P. and P. Molnar. Active deformation of Asia:from kinematics to dynamic [J]. Science,278:647-650,1997b

England P. and P. Molnar. The field of crustal velocity in Asia calculated from quaternary rates of slip on faults[J].Geophy.J.Int.,130:551-582,1997a

Haines, A. J., J. A. Jackson, W. E. Holt, and D. C. Agnew, Representing distributed deformation by continuous velocity fields, Sci. Rep. 98/5, 100pp., Inst. Of Geol. And Nucl. Sci.,

Lower Hutt, New Zealnad, 1998

Holt, et al., Earthquake strain rates and instantaneous relative motions within central and Eastern Asia[J], Geophys. J. Int., 122, 569-593, 1995

Hong Hanjing ,Wang Yipeng, et al. Mean velocity field and its dynamic implication of continental crust-Block's motion in China[A],Research on active fault [C], Beijing: Seismological Press（6）:17-29,1998.

http://www.seismicity.Harvard.edu.com.

Hu M.C. ,Lu Fu, Modern Geodesy, Beijing:Publishing House of Surveying and Mapping,1994.

Li Yanxing, Yang Guohua ,Li Zhi *et al.*, Movement and strain state of China active blocks, *Science in China (D)*, 33(Supplement): 65-81, 2003

Liu Jingnan,Shi Chuang,Xu Caijun, Jiang Weiping, Present-day crustal movement speed of China continent block using local repeated GPS network [J].Geomatics And Information Science Of Wuhan University, .26(3): 189-195,2001

Ren Jinwei, Deformation kinematics of Tibetan Plateau determined from GPS observations , Study on the China mainland main active zones present-day crustal motion and geodynamics[M],Lai Xi'an , Liu Jingnan , Beijing: Seismological Press，342-355,2001

Wang M., Shen Z., Niu Z., *et al.*, Contemporary crustal deformation of Chinese continent and tectonic block model, *Science in China (D)*, 33(Supplement):21-33, 2003

Wang Q., Zhang P., J. Freymueller, *et al.*, Present-day crustal deformation in China constrained by Global Positioning System measurements, *Science,* 294, 574-577, 2001

Xu Caijun, Wang Qi, Liu Jingnan, *et al.*, Recent Crustal Deformation And Strain Accumulations In Continental China Inferred From GPS And Seismicity Data. In: Adam and Schwarz (eds.), Vistas for Geodesy in the New Millennium. International associate of geodesy symposia, 125, Berlin: Springer,512-517, 2002

Xu Xiwei, Cheng Guoliang ,Ma Xingyuan et al., Rotation model and dynamics of blocks in North China and its adjacent areas, Earth Science-Journal of China University of Geosciences, 19(2):129-138,1994.

Zhang Peizhen, Late quaternary tectonic deformation and earthquake hazard in continental China, QUATERNARY SCIENCES 5:404- 413,1999

Crustal deformations at permanent GPS sites in Denmark

Shfaqat Abbas Khan[1], Per Knudsen[1] and Carl Christian Tscherning[2]

[1] National Survey and Cadastre, Geodetic department, Rentemestervej 8, 2400 Copenhagen NV, Denmark. e-mail: sak@kms.dk
[2] Dept. of Geophysics, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark. e-mail: cct@gfy.ku.dk

Abstract. The National Survey and Cadastre (KMS) is responsible for the geodetic definition of the reference network in Denmark. Permanent GPS stations play an important role in the monitoring and maintenance of the geodetic network. During 1998 and 1999 KMS established three permanent GPS station in Denmark, SMID, SULD and BUDP. Using almost 4.5 years of continuous data from the Danish station and the Swedish station, ONSA, we analyse the daily GPS solution due to crustal deformation caused by glacial isostatic adjustment (GIA). Although, displacements due to GIA are only 1-3 mm/year at the Danish GPS sites, the current precision of positioning using GPS allows us to observe these effects. The modelled horizontal GIA velocities and the observed horizontal residuals obtained from GPS show almost the same direction for all station. However, the observed velocity residuals are larger than the modelled GIA velocities. Furthermore, the difference of the observed vertical velocity residual rate between ONSA and SMID, and between ONSA and SULD seems to agree with the expected rate according to the postglacial model by Milne. However, the vertical velocity rate of BUDP seems to disagree with the model. Moreover, the analysis presented here uses a relative short period of data. The error of the vertical velocities (including reference frame drift) is order of ± 1.1 mm/yr.

Key words. GPS – Glacial Isostatic Adjustment

1 Introduction

Glaciers and ice sheets have a strong influence on the physical shape and environment of the earth. As they melt or grow during the climatic changes, they deform the Earth resulting in isostatic adjustment. The recover of the isostatic balance may last many thousand years depending on the ice thickness during the Temperature Minimum. At present time Glacier ice covers 1/10 of the earth's land surface, however during an ice age the ice may cover 1/3 of the land surface. The last ice age, also know as 'Late Glacial Temperature Minimum' occurred nearly 23.000 years ago and covered Canada, Greenland, Fennoscandia and Siberia in the north and the Antarctica in the south. During LGTM the ice sheet of Fennoscandia reached a thickness of more than 3 km resulting in a long term isostatic adjustment.

The present deformations due to the LGTM in Fennoscandia have rates of more than 10 mm/year in the vertical direction in the northern Scandinavia. In the the horizontal direction the displacements are smaller, about 1-3 mm/year. In Denmark, the postglacial rebound is relatively small in both the vertical and the horizontal direction. Figure 3 displays vertical and horizontal velocities obtained using the Milne-model (Milne *et al.*, 2001). According to the model, the ice sheet covered northern Denmark (see the zero line) while the southern area remained free from ice. In this study we compare the model deformation (based on the climatic history) with point displacements observed by GPS.

Although, crustal displacements at the Danish GPS sites are 1-2 mm/year, the current precision of positioning using GPS allows us to observe these effects. However, in order to achieve the millimetre accuracy of point positioning, several effects have to be considered e.g. precise satellite orbit determination, corrections due to tidal effects and modelling of the troposphere. Several previous studies have compared GPS measurements with modelled postglacial adjustment. Scherneck *et al.* (2002) and Johansson *et al.* (2002) compare measurements from continuously operating GPS receivers in Sweden and Finland with predicted calculations of the glacial isostatic adjustment (GIA) in Fennoscandia. Furthermore, Wahr *et al.* (2001) compare measurements from two GPS sites in Greenland with a GIA model. However, here we will focus on the Danish sites. The aim of this study is to include the Danish sites into the ongoing measurements of the GIA in Fennoscandia.

2 GPS Data Analysis

The National Survey and Cadastre (KMS) is responsible for the geodetic definition of the reference network in Denmark. Permanent GPS stations play an important role in the monitoring and maintenance of the geodetic network. Furthermore, KMS supports the international GPS infrastructure and research by supporting the International GPS Service (IGS). In September 1998 KMS established two permanent GPS stations, one in Suldrup (SULD) and one in Smidstrup (SMID).

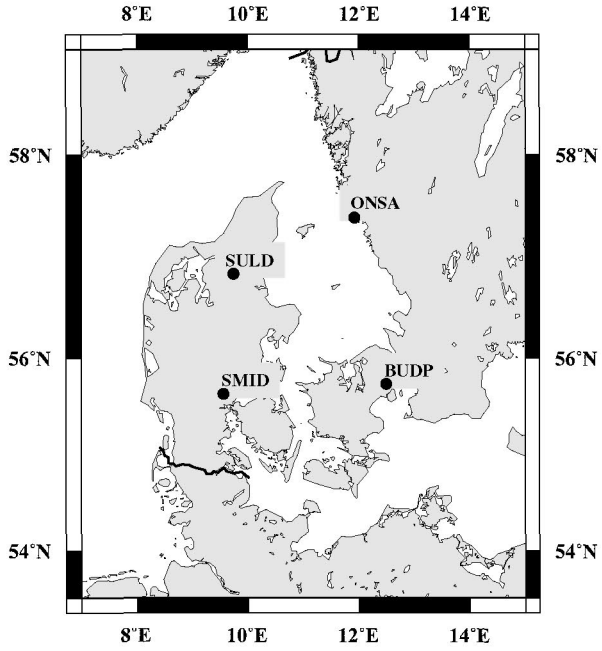


Fig. 1. Map of Denmark showing the location of the GPS stations.

Besides these two stations a third station, BUDP, was established in Buddinge in January 1999. All three stations are mounted on solid concrete pillars, which are electrically heated to prevent thermal expansion. The data is of high quality and useful in many geophysical studies, e.g. continental drift, atmosphere loading or measurements of the postglacial rebound. Here, we use data measured during the period from January 1999 to April 2003. Additionally, we use data from the Swedish permanent GPS station ONSA in our study. The locations of the GPS stations are shown in Figure 1.

We have analysed the GPS data using the GIPSY OASIS II software (Zumberge *et al.*, 1997) developed at the Jet Propulsion Laboratory (JPL). We use GPS orbits, earth orientation, and clock products obtained by JPL based on a global network of GPS sites. Receiver clock parameters and atmosphere parameters were modelled and a cut-off angle of 15 degrees was used. The data were corrected due to the solid earth tide and the ocean tide loading effect. The amplitudes and phase lags of the main ocean tide loading signal were calculated by the online loading provider by H.-G. Scherneck, www.oso.chalmers.se/~loading

Using the precise point positioning approach, we computed site coordinates in the nonfiducial frame and then transformed the coordinates into the ITRF00 frame. The transformation parameters from the free-network estimates to ITRF00 for a particular day are available at the following ftp site, sideshow.jpl.nasa.gov/pub/gipsy_products/x-files

The velocity vector, $\boldsymbol{V}$, describing the rigid plate motion is given by (Nocquet *et al.*, 2001)

$$\boldsymbol{V} = \Omega \boldsymbol{X} \tag{1}$$

where $\boldsymbol{X}$ is the site coordinate vector and Ω is the Euler matrix,

$$\Omega = \begin{pmatrix} 0 & -w_z & w_y \\ w_z & 0 & -w_x \\ -w_y & w_x & 0 \end{pmatrix}$$

Furthermore, the rotation parameters used in our study are $(w_x, w_y, w_z) =$
(0.081 ± 0.021, 0.489 ± 0.008, -0.792 ± 0.026) given in mas/year. These parameter are adopted from Kierulf *et al.* (2003). In addition, postglacial rebound is taken into account in the estimation of these parameters. Hence, postglacial rebound is not absorbed in the estimation of the rigid plate rotation parameters (Kierulf *et al.*, 2003) .

3 Results

Daily solutions of site positions obtained using the GIPSY OASIS II software are shown in Figure 2. The mean value of the particular component (north, east and vertical) is removed. The left figures show the north-south component (counted positive in the north direction). The figures in the middle show the east-west component (counted positive in the east direction) and the right figures show the vertical component (counted positive upward). The displacements are given in mm. From top to bottom, we display time series of the four GPS sites, BUDP, SMID, SULD and ONSA. The solid light lines show a linear fit (using least square adjustment) to the daily point solutions. The rigid motion of Eurasia is significant and may be removed using site velocities calculated by equation 1. The estimated velocity residuals after removing the rigid plate rotation are shown in Table 1.

In order to estimate the error, we use the 1σ formal errors generated by the GIPSY OASIS II software. Furthermore, we assume the errors are uncorrelated with respect to time. Errors in the vertical due to reference frame drift are simply estimated by assuming that the drift error is given as (see Argus and Heflin , 1995).
$\sigma_{drift} = 5/(\text{number of years})$
Thus, we get
$\sigma = \sqrt{{\sigma_{drift}}^2 + {\sigma_{GIPSY}}^2}$
Similar the horizontal errors are obtained by adding (in quadrature) the error of the rigid plate rotation (see previous section). The last column in Table 1 shows the modelled glacial isostatic adjustment by Milne *et al.* (2001). The model is based on the ICE-3G ice model and the PREM earth model with a lithosphere thickness of 120 km. For all the stations, the observed velocity residuals are larger than the modelled glacial isostatic adjustment. This is illustrated in Figure 3. The solid lines displays contour lines of the vertical velocities rates according to the GIA model by Milne and the red coloured vectors display the horizontal velocity rates due to the Milne GIA model. The blue vectors display the observed horizontal velocity residual. The modelled horizontal GIA velocities and the observed horizontal residuals show almost the same direction for all station. However, the observed velocity residuals are larger than the modelled velocities at all station. This may indicate that the modelled GIA velocities

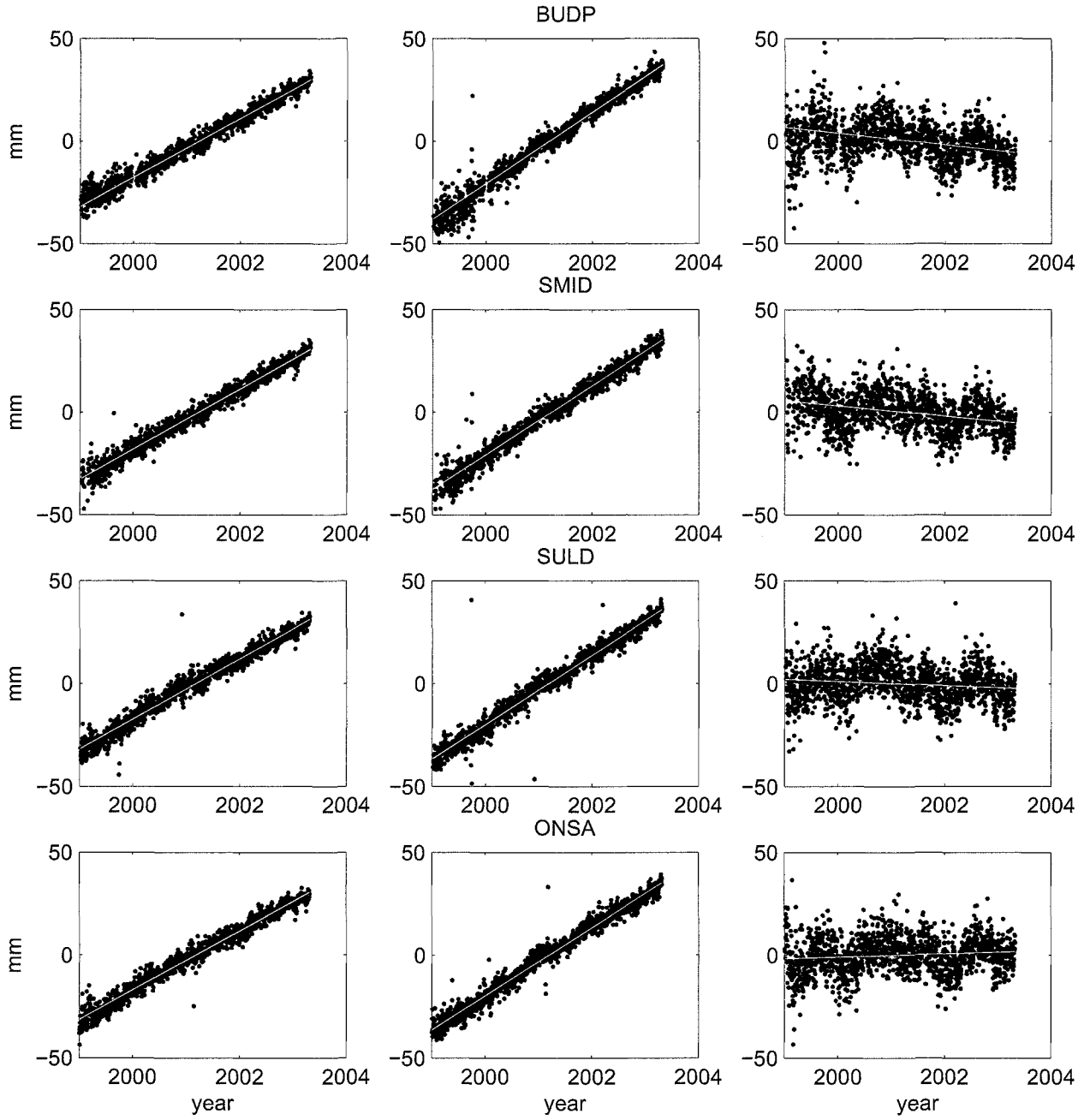


Fig. 2. Time series of site positions obtained using the GIPSY OASIS II software. The mean value of each component (north, east and vertical) is subtracted. The left figures show the north component, the figures in the middle show the east component and the right figures show the vertical component (all components are given in mm). From top to bottom, we display the four GPS sites, BUDP, SMID, SULD and ONSA. The solid light lines show a linear fit to the observations. .

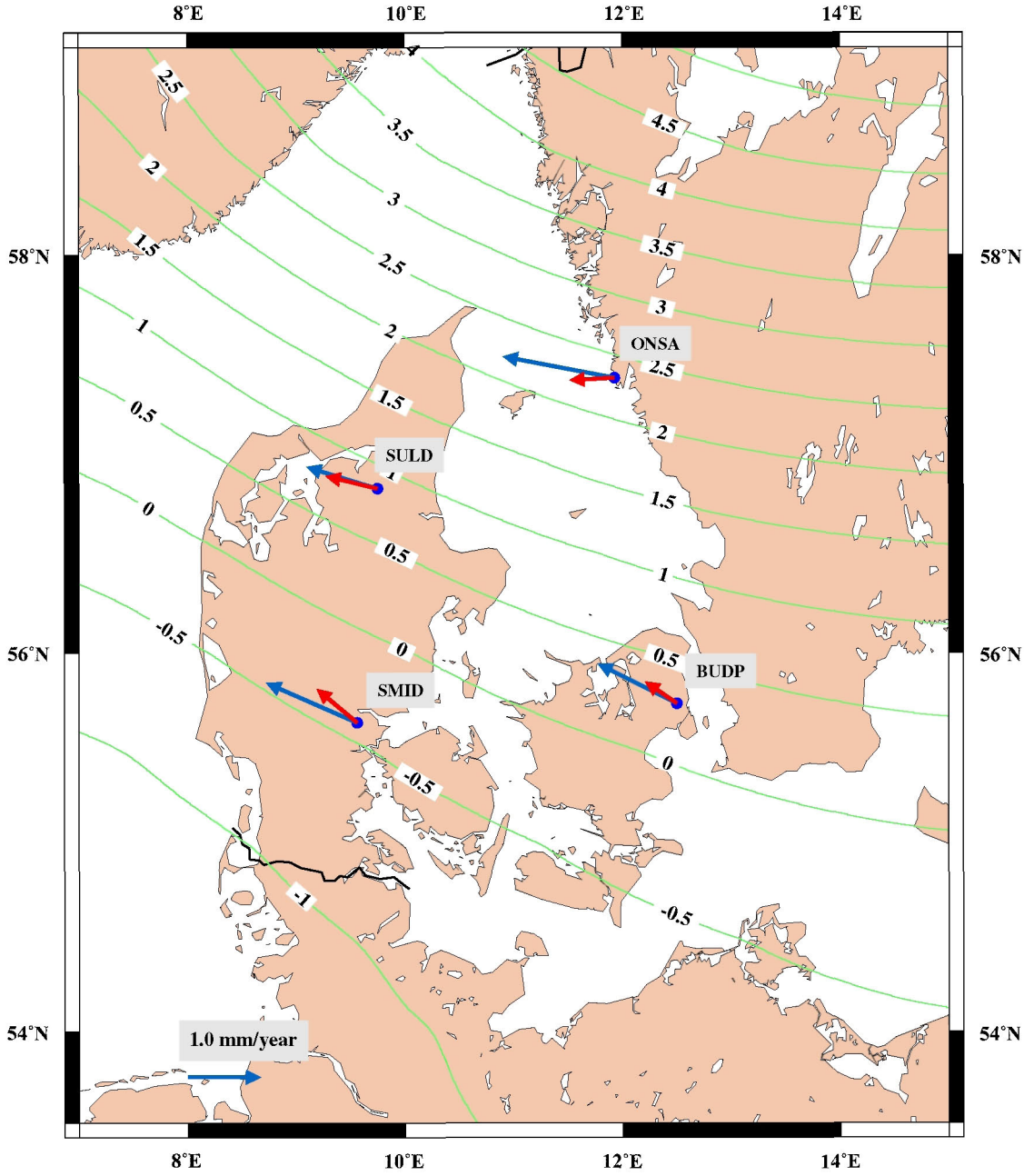


Fig. 3. The solid lines displays contours [with interval of 0.5 mm/year] of the verticalvelocitiesrates based upon the GIA model by Milne and the red collared vectors display the horizontal velocity rates due to the Milne model. The blue vectors display the observed horizontal velocity residual.

are too small by a factor, or perhaps the rigid plate motion is not fully removed.

Daily point solutions of the vertical displacements obtained by the GIPSY OASIS II software (displayed in figure 2, right) can be better determined if one of the stations are kept fixed. However, then we have to consider relative deformations between the sites. In addition, the rate of the vertical velocity residual is very much dependent on which reference frame is used. Hence, using e.g. the ITRF97 we would obtain different results from those listed in Table 1. Therefore, we find it more appropriate to compare differences in the vertical velocities between the involved stations rather than the absolute values. The difference of the observed vertical velocity residual rate between ONSA and SMID seems to agree with the expected rate according to the Milne GIA model. Furthermore, the observed residual rate between ONSA and SULD seems to be consistent with the model rate. However, when considering the difference of the vertical rate between ONSA

Table 1. GPS site coordinates.

Station name		Observed residual [mm/year]	standard deviation [mm/year]	predicted GIA (Milne) [mm/year]
BUDP	North	0.54	0.36	0.30
	East	-1.09	0.42	-0.44
	Vertical	-3.58	1.13	0.24
SMID	North	0.55	0.35	0.47
	East	-1.26	0.42	-0.56
	Vertical	-2.66	1.12	-0.45
SULD	North	0.31	0.35	0.18
	East	-1.00	0.41	-0.71
	Vertical	-1.07	1.11	0.80
ONSA	North	0.29	0.35	-0.03
	East	-1.54	0.41	-0.62
	Vertical	0.70	1.11	2.26

and BUDP, there seems to be an disagreement between the observed and the modelled rates. BUDP is moving downward with a velocity rate, which is not explainable with the GIA only.

4 Conclusion

Data from three Danish and one Swedish continuously operating GPS stations have been analysed due to vertical and horizontal displacements. The modelled horizontal GIA velocities and the observed horizontal residuals show almost the same direction for all station. However, the observed velocity residuals are larger than the modelled GIA velocities. The reason for this disagreement could be that the rigid plate motion is not fully removed, or that the predicted GIA is not fully correct. However, this can be investigated using more GPS stations and by considering a larger area. The spatial distribution of the two effects is different. Hence, by considering a larger area it is possible to separate these effects. Furthermore, The difference of the observed vertical velocity residual rate between ONSA and SMID seems to agree with the expected rate according to the Milne GIA model. The observed residual rate between ONSA and SULD are also consistent with the model rate. However, BUDP is moving downward relative to ONSA by a velocity, which is almost twice the expected rate due to the GIA model.

Moreover, here we use almost 4 1/2 years of continues GPS data for each site. The sites are still operating and more data will contribute significantly for a better determination of the vertical component. In addition KMS are monitoring 3 new stations, which will be in operation in the beginning of 2004.

Acknowledgements. We thank Hans-Georg Scherneck, Jan M. Johansson and Martin Lidberg for their support throughout this study. Additional information regarding uplift due to the Fennoscandia ice sheet is available from the following web-sites:
http://www.oso.chalmers.se/~hgs

References

Argus, D.F and M.B. Heflin (1995) plate motion and crustal deformation estimated with geodetic data from the Global Positioning System, Geophys. Res. Lett., 22, 1973-1976.

Johansson, J.M., J.L. Davis, H.-G. Scherneck, G.A. Milne, M. Vermeer, J.X. Mitrovica, R.A. Bennett, G. Elgered, P. Elsegui, H. Koivula, M. Poutanen, B.O. Rnnng, and I.I. Shapiro (2002) Continuous GPS measurements of postglacial adjustment in Fennoscandia, 1. Geodetic Results, J. Geophys. Res., Vol. 107, No. B8, p. 10,1029.

Kierulf, H.P., H.-P. Plag, and O. Kristiansen (2003) Towards the true rotation of a rigid Eurasia, EUREF Publication No. 11, Verlag des Bundesamtes für Kartographie undGeodäsie, Frankfurt am Main.

Milne, G.M., J. L. Davis, Jerry X. Mitrovica, H.-G. Scherneck, J. M. Johansson, M. Vermeer, H. Koivula (2001) Space-Geodetic Constraints on Glacial Isostatic Adjustment in Fennoscandia, Science, Vol. 291.

Nocquet, J.-M., E. Calais, Z. Altamimi, P. Sillard, C. Boucher (2001) Intraplate deformation in western Europe deduced from an analysis of the International Terrestrial Reference Frame 1997 (ITRF97) velocity fi eld, J. Geophys. Res., Vol. 106 , No. B6 , p. 11,239.

Scherneck, H.-G., J.M. Johansson, G. Elgered, J.L. Davis, B. Jonsson, G. Hedling, H. Koivula, M. Ollikainen, M. Poutanen, M. Vermeer, J.X. Mitrovica, and G.A. Milne (2002) BIFROST: Observing the Three-Dimensional Deformation of Fennoscandia in: Ice Sheets, Sea Level and the Dynamic Earth, Mitrovica, J.X. and B.L.A. Vermeersen (eds), pp. 69-93, AGU Geodynamics Series, Vol. 29, American Geophysical Union, Washington D.C.

Wahr, J.M., T. van Dam, K. Larson, and O. Francis (2001) Geodetic Measurements in Greenland and their implications, J. of Geophys. Res., 106, p. 16567 - 16581.

Zumberge, J.F., M.B. Heflin, D.C. Jefferson, M.M. Watkins and F.H. Webb (1997) Precise point positioning for the effi cient and robust analysis and GPS data from large networks, J. Geophys. Res., 102, p. 5005-5017.

Monitoring Land Subsidence of Jakarta (Indonesia) Using Leveling, GPS Survey and InSAR Techniques

Hasanuddin Z. Abidin, H. Andreas, M. Gamal
Department of Geodetic Engineering, Institut Teknologi Bandung (ITB)
Jl. Ganesha 10, Bandung 40132, INDONESIA, E-mail : hzabidin@gd.itb.ac.id

Rochman Djaja, C. Subarya
National Coordinating Agency for Survey and Mapping, Cibinong, Bogor, INDONESIA

K. Hirose,
Nikko Exploration & Development (NED) Co., Tokyo, JAPAN

Y. Maruyama
Earth Remote Sensing Analysis Center (ERSDAC), Tokyo, JAPAN

D. Murdohardono, H. Rajiyowiryono
Directorate of Environmental Geology and Mining Area Management, Bandung, INDONESIA

Abstract. Jakarta is the capital city of Indonesia with a population of about 12 million people, inhabiting an area of about 25-km by 25-km. It has been reported for quite sometime that several locations in Jakarta are subsiding at different rates. Leveling surveys performed in 1982, 1991 and 1997 have detected the subsidence up to about 80 cm during the period of 1982-1991, and up to about 160 cm during the 1991-1997 period; while GPS surveys observed the subsidence up to about 50 cm during the period of 1997-2002. InSAR technique using JERS-1/SAR L-band data estimated a subsidence rate of about 5 to 10cm during the period of 1993 to 1995. Maximum subsidence were found in the northwestern and central eastern parts of Jakarta, while minimum subsidence were found in the southern part. InSAR results show good correspondence with the results from Leveling and GPS Surveys

Keywords. Land subsidence, Jakarta, Leveling, GPS Survey, InSAR.

1 Introduction

Jakarta is the capital city of the Republic of Indonesia with a population of about 12 million people, inhabiting an area of about 650.40 km^2. Jakarta is centered at the coordinates of about $6^0 15'$ S (latitude) and $106^0 50'$ E (longitude) and, located on the lowland of the northern coast of the West Java province, Indonesia The area is relatively flat, with the topographical slopes ranging between 0 and 2 degrees in the northern and central parts, and between 0 and 5 degrees in the southern part. The southern-most area of Jakarta has an altitude of about 50-m above mean sea level.

Land subsidence is not a new phenomenon for Jakarta. It has been reported for many years that several places in Jakarta are subsiding at different rates. According to the Local Mines Agency of Jakarta, over the period of 1982 to 1997, subsidence ranging from 20 cm to 200 cm is evident in several places in Jakarta. The occurrence of land subsidence in Jakarta was realized for the first time in 1926. Evidence for subsidence was based on repeated leveling measurements conducted in the northern part of Jakarta. Unfortunately this investigation of land subsidence using leveling had not been continued for 50 years until 1978. Starting in 1978, the impact of land subsidence in Jakarta could be seen in several forms, such as the cracking of permanent constructions located around the center of the Jakarta area (along Thamrin street), the wider expansion of flooding areas, the lowering of the ground water level, and increased inland sea water intrusion.

Since the early 1980's, the land subsidence in several places of Jakarta has been measured using several techniques, e.g. leveling surveys, extensometer measurements, ground water level observations, GPS (Global Positioning System) surveys, and InSAR (Interferometric Synthetic Aperture Radar) techniques. The prediction of ground subsidence, based on certain models incorporating geological and hydrological parameters of Jakarta, has also been investigated by a few researchers [*Murdohardono and*

Tirtomihardjo, 1993; *Yong et al.*, 1995; *Purnomo et al.*, 1999]. This paper concentrates on the use of leveling, GPS surveys and InSAR for monitoring the land subsidence in Jakarta.

2 Land Subsidence in Jakarta

According to *Murdohardono & Sudarsono* (1998) and *Rismianto & Mak* (1993), there are four different types of land subsidence that can be expected to occur in the Jakarta basin, namely: subsidence due to groundwater extraction, subsidence induced by the load of constructions (i.e. settlement of high compressibility soil), subsidence caused by natural consolidation of alluvium soil, and geotectonic subsidence. From those types of subsidence, the main spectrum of land subsidence in Jakarta is thought to be caused by groundwater extraction. Excessive groundwater extraction will lead to the deepening of groundwater level (piezometric head), which in turn will cause land subsidence and also seawater intrusion.

Three aquifers are recognized inside a 250 m thick sequence of quaternary sediment of the Jakarta basin, namely [*Hadipurwo*, 1999] : the Upper Aquifer, an unconfined aquifer, occurs at a depth of less than 40 m; the Middle Aquifer, a confined aquifer, occurs at a depth between 40 and 140 m; and the Lower Aquifer, a confined aquifer, occurs at a depth between 140 and 250 m. Inside those aquifers, the groundwater generally flows from south to the north. Below a depth of 250 m, an aquifer in the tertiary sediment was also found. But according to *Murdohardono and Tirtomihardjo* (1993), it is less productive and its water quality is relatively poor.

The groundwater extraction in Jakarta could be categorized into shallow (< 40 m) and deep (> 40 m) extraction. Shallow extraction is through dug wells or driven wells, operated with buckets, hand pumps or small electrical pumps; whereas the deep extraction is mostly from drilled wells. Shallow extraction is mostly done by the population. It is well spread over the area, but its extraction rate per well is relatively low. Deep extraction is usually conducted by industry. It is usually more concentrated, and has a relatively high extraction rate per well. The land subsidence observed in the coastal, west, and northeastern parts of Jakarta are thought to be caused by this deep groundwater extraction, which reduces the water pressure in the aquifer (piezometric level) [*Rismianto & Mak*, 1993].

According to *Soetrisno, et al.* (1997), the piezometric level in North Jakarta has changed from 12.5 m above sea level in 1910 to about sea level in 1970's, and then deepened significantly to 30-50 m below sea level in 1990's.

It should be noted that in the case of Jakarta, the comprehensive information on the characteristics of land subsidence would be important for several tasks, such as spatial-based groundwater extraction regulation, effective control of flood and seawater intrusion, conservation of environment, design and construction of infrastructures, and spatial development planning in general.

3 Land Subsidence Measured by Leveling Surveys

The establishment of vertical control in Jakarta was started in 1925 at the time of Dutch colonization by using optical leveling measurements. The first precise leveling network was established between May 1925 and April 1926. This network has a leveling line of about 38 km. Unfortunately the data and results of these leveling surveys are lost and unknown at the present time. After these first surveys, the next systematic leveling surveys covering Jakarta area were conducted in 1978, 1982, 1991, 1993, and 1997. The leveling surveys were done using Wild N3, Zeiss Ni002, and Wild NAK precise levels. Each leveling line was measured in double-standing mode, and each leveling session was measured forward and backward. The leveling line for each session is about 1 km in length. The tolerance for the difference between the forward and backward height-difference measurements is set to be $4\sqrt{D}$ mm, where D is the length of leveling line in km.

The height differences are computed based on parametric least squares adjustment. Prior to adjustment, the data are corrected by rod constant and refraction corrections. After checking the number of collocated points between the surveys and data quality, only three surveys were considered in this case, i.e. those conducted in 1982, 1991, and 1997. Moreover, only the results from 45 leveling points in the network, which are considered the most reliable, are used. The distribution of these points is shown in Figure 1.

It should be noted here that the leveling reference point of Jakarta is the bench mark of Tanjung Priok tide station located in the central northern part of Jakarta. Since this area is also a subsidence prone area, stability of this reference point is questionable. No systematic study has been done yet to investigate its stability; and since it can not be occupied by GPS receiver, this

point was not included in the GPS land subsidence monitoring network of Jakarta. In analyzing the leveling derived subsidence, the reference point was assumed to be stable.

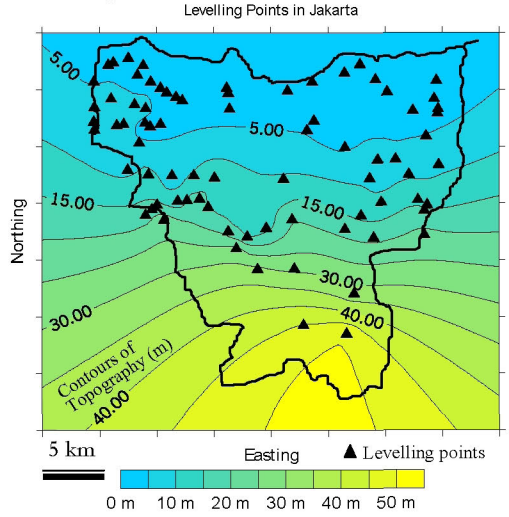


Fig.1 Leveling network in Jakarta

It was obtained from leveling surveys that a maximum land subsidence observed during the period of 1991-97 is about 160 cm, while for the period of 1982-91 is about 80 cm. Total subsidence in the period of 1982 to 1997 is shown in Figure 2.

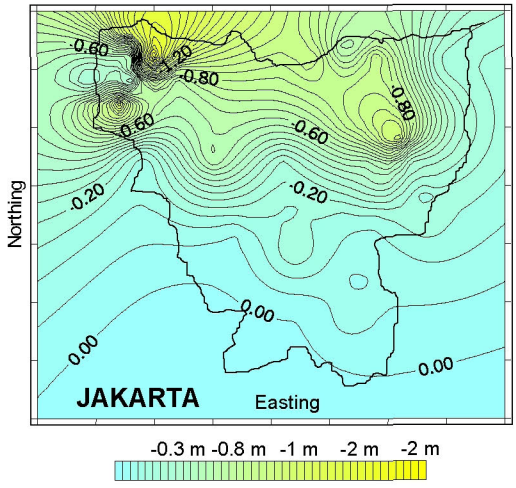


Fig. 2 Leveling derived land subsidence in Jakarta (1982 – 1997 period).

During the period between 1982 and 1991, the maximum rate of subsidence is about 8 cm/year, while during the period between 1991 and 1997 it is about 26 cm/year. But in general the subsidence rate is about 1-6 cm per year as shown by the box-and-whisker plots in Figure 3.

A box-and-whisker plot is a histogram-like method of displaying data, which its parameters are shown in Figure 4.

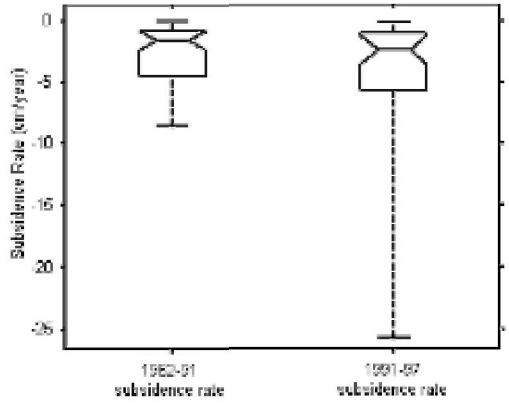


Fig. 3 Box-and-Whisker plots of land subsidence rates in Jakarta.

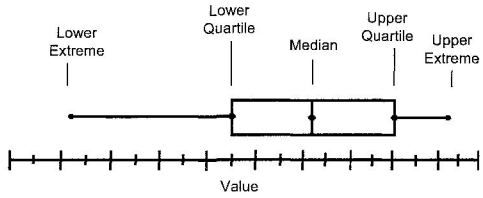


Fig. 4 Parameters of Box-and-Whisker plot.

The results on Figures 2 and 3 show that land subsidence in Jakarta has both spatial and temporal variations. The subsidence rate was found to be closely related to the rate of piezometric water level deepening in the middle and lower aquifers caused by over discharging of groundwater [*Abidin et al.,* 2001]. More detail analysis of these leveling results in relation with the hydro-geological data and information of Jakarta basin can be seen in *Abidin et al.* (2001).

4 Land Subsidence Measured by GPS Surveys

The use of GPS satellite-based positioning system to systematically establish the geodetic control points all over Jakarta was firstly conducted in 1994 by the National Land Agency (BPN). This GPS network is aimed at supporting cadastral mapping in the Jakarta area and its design was not intended for monitoring the land subsidence in Jakarta. Considering the higher efficiency and effectiveness of GPS surveys compared to the leveling surveys in monitoring the land subsidence of Jakarta, the Department of Geodetic Engineering, Institute of Technology Bandung (ITB) decided to establish the new GPS

network for monitoring land subsidence in Jakarta basin, where some of its points are also the points of the existing BPN network. The configuration of this GPS monitoring network at the present time is shown in Figure 5. BAKO, the southern most point in the network and also the Indonesian zero order geodetic point, is considered as a stable reference point. BAKO is an IGS station, operated by the National Coordinating Agency for Survey and Mapping (BAKOSURTANAL).

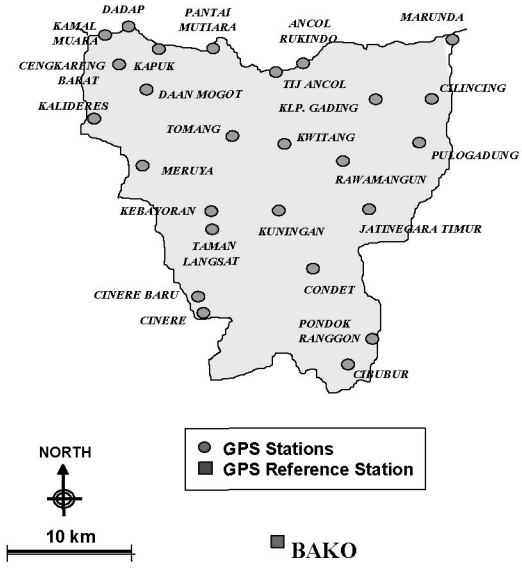


Fig. 5 Distribution of GPS stations for monitoring the land subsidence in Jakarta

Seven GPS surveys have been conducted, namely on the periods of 24 - 26 Dec. 1997, 29 - 30 June 1999, 31 May - 3 June 2000, 14 - 19 June 2001, 26 - 31 Oct. 2001, 02 - 07 July 2002, and 21 - 26 Dec. 2002. The GPS surveys at all stations were all carried out using dual-frequency geodetic-type GPS receivers. For GPS surveys, the length of sessions was between 12 to 24 hours, respectively. The data were collected with a 30 seconds interval, and elevation mask was set at 15^0 from all stations. The data of GPS surveys was processed using the scientific software Bernesse 4.2 [*Beutler et al.,* 2001]. Since we are mostly interested with the relative height component of the coordinates with respect to a stable point, the radial processing mode was used instead of network adjustment mode. In this case the relative ellipsoidal heights of all stations are determined relative to BAKO station, which is assumed to be a relatively stable point.

Considering the length of the baselines, which could be up to 40 - 50 km, a precise ephemeris was used instead of the broadcast ephemeris for data processing. The effects of tropospheric and ionospheric biases are mainly reduced by the differencing process. The parameters of residual tropospheric bias for individual stations are then estimated to further reduce the tropospheric effects. In the case of the residual ionospheric delay reduction, a local ionospheric modeling is implemented. The algorithms for these tropospheric parameter estimation and local ionospheric modeling could be seen in *Beutler et al.* (2001). In processing baselines, most of cycle ambiguities of the phase observations were successfully resolved.

Standard deviations of GPS derived relative ellipsoidal heights from all surveys were in general better than 1 cm (see Figure 6). A few points have slightly worse standard deviations due to the lack of observed data caused by the signal obstruction by trees and/or building around the station.

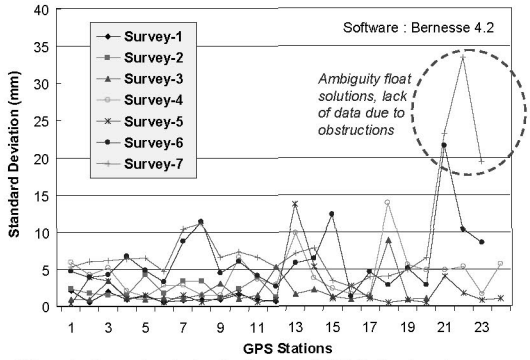


Fig. 6 Standard deviations of GPS derived relative ellipsoidal heights in mm.

Land subsidences derived from GPS surveys in summary are shown in Figures 7 and 8. In general the estimated subsidence rates are around 1 to 10 cm per year, depending on the location. In comparison with previous rates obtained from leveling results, it can be concluded that land subsidence in Jakarta is still continuing with a mean rate of about 5 to 7 cm per year.

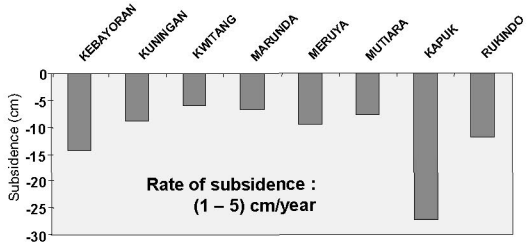


Fig. 7 GPS derived land subsidence (cm) in the period of Dec. 1997 to Dec. 2002.

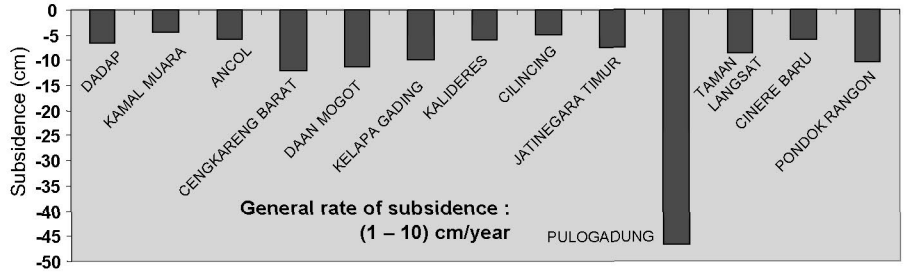


Fig. 8 GPS derived land subsidence (cm) in the period of June 2001 to Dec. 2002.

5 Land Subsidence Measured by InSAR Technique

The principles of InSAR for estimating changes in the Earth's surface are explained in *Massonnet & Feigl* (1998). In this study land subsidence is estimated by differential InSAR processing using JERS-1 SAR data. The principal specification of JERS-1/SAR is L-band (23.5cm wavelength), HH polarization and about 35 degrees of the off-nadir angle.

JERS-1/SAR in total acquired 17 scenes from Jakarta area during the period of 1993/02/25-1998/09/11. For InSAR processing, 41 image pairs were selected and co-registered. The VEXCEL 3D software was used for processing. Data processing was started from single look complex (SLC), which then continued by resampling SLC (co-registration), creating interferogram, filtering, unwrapping interferogram, and refining process. The size of multi-looking was 2 pixels in range by 6 lines in azimuth.

Three-pass differential InSAR technique is used in this study. In order to remove topographical fringe from the primary InSAR fringe, DEM is generated by using the 44-day interferogram. An external DEM was not used since it was not publicly available. If longer pair data is applied to remove topographical fringe, it will be impossible due to no coherence caused by strong subsidence. Actually even 44 days is thought as longer pair. But there was no pair less than 44 days interval found in JERS-1/SAR data. In this study only four pairs with relatively good coherences were found out for different time periods as shown in Table 1.

The differential interferograms were generated by above data sets and displacement maps were obtained. In this case the line-of-sight displacement values were converted to vertical displacements, and flattening was automatically carried out by using information of the satellite orbital data and DEM generated by SAR. Two subsidence maps are shown in Figure 9.

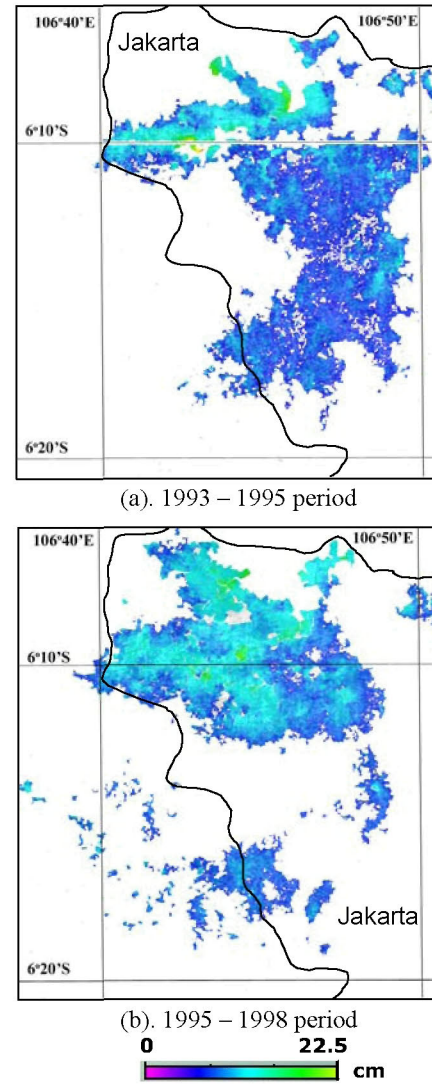


Fig. 9 InSAR derived land subsidence (cm) in the middle and north western parts of Jakarta.

Figure 9 shows that land subsidence are seen in the north and northwest part of the images. The same areas of subsidence are shown in the images derived from different data sets, which

Table 1. The pair images used in this study.

Period (Date)	Baseline (m)	Interval (days)
1993/10/03-1995/09/07	351	704
1995/09/07-1995/10/21	707	44
1995/10/21-1998/09/15	641	1056
1997/01/03-1997/05/15	87	132

are showing the same trend as NE-SW direction. It is corresponding to the results from leveling survey. The annual rate of land subsidence are estimated approximately 10 cm/year (1993-1995) and 6 cm/year (1995-1998). These figures of rates are supported by the results obtained from leveling and GPS surveys.

6. Closing Remarks

This study shows that combination of leveling, GPS surveys, and InSAR results are useful for studying and monitoring land subsidence phenomena. Besides complementing each other, both spatially and temporally, they can also check against one another for quality assurance purposes. Leveling can flexibly handle relatively dense and crowded urban areas; GPS is good for relatively open areas; and InSAR is relatively not good for vegetated areas and other areas with relatively rapid environmental changes. In case of Jakarta, the combination of those methods will be useful in studying spatial and temporal characteristics of land subsidence phenomena.

In order to have more insight into the land subsidence process, the above geodetic results should also be integrated and correlated with the hydro-geological data and information. Considering the importance of land subsidence data and information, monitoring land subsidence in Jakarta basin will be continued. Next GPS survey will be done at the end of 2004. In the case of using InSAR techniques, processing the images from other satellites (e.g. ERS-1/2, RADARSAT and Envisat) will be attempted in the near future. It will also be encouraging to wait for the launch of PALSAR, which is successor of JERS-1 and have function of L-band multi polarization. L-band data is less susceptible to loss in coherence even though repeat passes period is longer; and multi polarization data is allowing a more sophisticated physical interpretation of SAR interferograms [*Stebler et al*, 2002].

References

Abidin, H. Z., R. Djaja, D. Darmawan, S. Hadi, A. Akbar, H. Rajiyowiryono, Y. Sudibyo, I. Meilano, M. A. Kusuma, J. Kahar, C. Subarya (2001). "Land Subsidence of Jakarta (Indonesia) and its Geodetic-Based Monitoring System." *Natural Hazards*. Vol. 23, No. 2/3, March, pp. 365 - 387.

Beutler, G., H. Bock, E. Brockmann, R. Dach, P. Fridez, W. Gurtner, U. Hugentobler, D. Ineichen, J. Johnson, M. Meindl, L. Mervant, M. Rothacher, S. Schaer, T. Springer, R. Weber (2001). *Bernese GPS software version 4.2*. U. Hugentobler, S. Schaer, P. Fridez (Eds.), Astronomical Institute, University of Berne, 515 pp.

Hadipurwo,S. (1999). "Groundwater." *In COASTPLAN JAKARTA BAY PROJECT, Coastal Environmental Geology of the Jakarta Reclamation Project and Adjacent Areas*, CCOP COASTPLAN Case Study Report No. 2, Jakarta/Bangkok, pp. 39 - 49.

Murdohardono,D. and H. Tirtomihardjo (1993). "Penurunan tananh di Jakarta dan rencana pemantauannya." *Proceedings of the 22nd Annual Convention of the Indonesian Association of Geologists*, Bandung, 6-9 Dec., pp. 346 - 354.

Massonnet, D. and K.L. Feigl (1998). "Radar Interferometry and its Application to Changes in the Earth's Surface." *Reviews of Geophysics*, Vol. 36, No. 4, November, pp. 441-500.

Murdohardono,D. and U. Sudarsono (1998). "Land subsidence monitoring system in Jakarta." *Proceedings of Symposium on Japan-Indonesia IDNDR Project : Volcanology, Tectonics, Flood and Sediment Hazards*, Bandung, 21-23 September, pp. 243 - 256.

Purnomo, H., D. Murdohardono, and H. Pindratno (1999). "Land Subsidence Study in Jakarta." *Proceedings of Indonesian Association of Geologists, Volume IV : Development in Engineering, Environment, and Numerical Geology*, Jakarta, 30 Nov.-1 Dec., pp. 53-72.

Rismianto,D. and W. Mak (1993). "Environmental aspects of groundwater extraction in DKI Jakarta: Changing views." *Proceedings of the 22nd Annual Convention of the Indonesian Association of Geologists*, Bandung, 6-9 Dec., pp. 327 - 345.

Stebler, O., E. Meier, D. Nuesch (2002). "Multi-baseline polarimetric SAR interferometry-first experimental spaceborne and airborne results". *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 56, pp. 149– 166

Soetrisno,S., Satrio H., and Haryadi T. (1997). "To Anticipate Impacts of Reclamation of Jakarta Bay, A Groundwater Conservation's Perspective." Paper presented at *Workshop on Coastal and Nearshore Geological/Oceanographical Assessment of Jakarta Bay : A Basis for Coastal Zone Management and Development*, Jakarta, 25 - 28 June.

Yong,R.N., E. Turcott, and H. Maathuis (1995). "Groundwater extraction-induced land subsidence prediction: Bangkok and Jakarta case studies." *Proceedings of the Fifth International Symposium on Land Subsidence*, IAHS Publication no. 234, October, pp. 89-97.

Microgravity survey in the southern tip of the Atera fault, Central Japan

T. Tanaka[1], H. Aoki[1], K. Oshita[2], K. Nozaki[2]
[1] Tono Research Institute of Earthquake Science, Association for the Development of Earthquake Prediction, Yamanouchi, Akeyo-cho, Mizunami 509-6132, Japan
e-mail: tanaka@tries.jp; Tel.: +81-572-67-3105; Fax: +81-572-67-3108
[2] OYO Corporation, Miyukigaoka, Tsukuba 305-0841, Japan

Abstract. The Atera fault is one of the most active both geomorphologically and geologically in Japan. In the southern part of the fault, the Sakashita body (Ueno volcano) erupted about 1.5 km southwest of the fault in a direction parallel to it (early Pleistocene). A local strike direction change, reversal of vertical displacement (SW uplift), and branching of the fault line occurred in the southern tip of the fault (in and around the Dendahara district). Our microgravity survey suggests that a high-density body (probably basaltic) prevents ordinary activity of the Atera fault (left-lateral and NE uplift).

Keywords. microgravity, active fault, volcano

1 Introduction

The Atera fault is one of the most famous active faults in Japan (Sugimura and Matsuda, 1965). There are many studies from both geomorphological and geological points of view. However, there are few geophysical surveys because of difficult access to the area around the fault (Fig. 1).

Tanaka et al. (2003) reported geophysical surveys of the Atera fault pursued by the Tono Research Institute of Earthquake Science (TRIES) and proposed that a group of vents of the Ueno volcano (the Sakashita body of the Ueno Basalts) is located about 1.5 km southwest of the Atera fault in a direction parallel to it. In this research, a supplement to Tanaka et al. (2003), we carried out a microgravity survey in the Dendahara district where the fault line intersects the southeastern extension of the Ueno volcano. Local change of strike of the fault, reversal of vertical displacement, and bifurcation of the fault have occurred in this area. Our aim is to clarify the origin of such peculiar structures.

2 Tectonic setting

Some trench studies have been done in the Dendahara district because of the clearness of fault topography (Toda et al., 1995). This area is the southern limit of the region where the Ueno Basalts have erupted onto the surface as a part of mantle

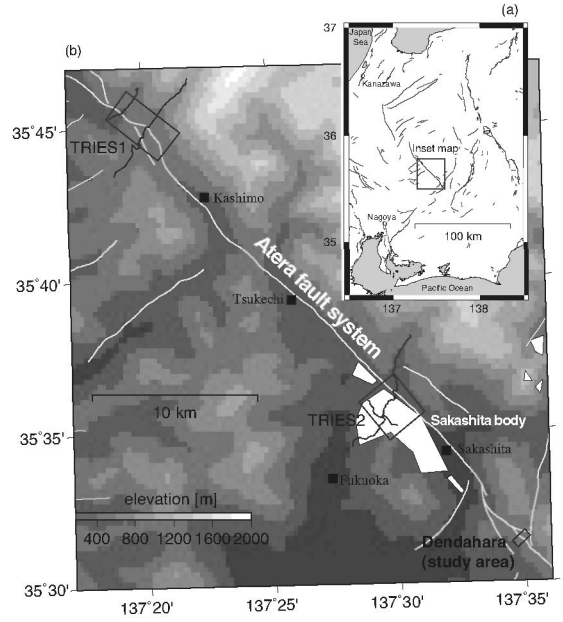


Fig. 1. Location map with active fault distribution (Nakata and Imaizumi, 2002) (a) The regional map shows active faults distribution in the Central Japan. (b) The topographic map around the Atera fault. Four black curves show seismic-receiver lines and three rectangulars show microgravity survey areas. White polygons show the Ueno basalt bodies. The largest one is called the Sakashita body (or Ueno volcano). The northern area (TRIES1) and the middle area (TRIES2) were discussed in TRIES(2002) and Tanaka et al. (2003), respectively. Tokyo datum is used as geodetic reference frame.

diapir (Kimura and Yoshida, 1999). Tanaka et al. (2003) discussed that the relationship between the Ueno volcano and the Atera fault. They pointed out the possibility that there was a group of vents of the Ueno volcano in about 1.5 km southwest of the Atera fault extending in a direction parallel to it by microgravity and reflection surveys with the constraint of well logging data by National Research Institute for Earth Science and Disaster Prevention (NIED). Therefore, the 'unoutcropped' Ueno volcano may geographically affect the growth of the Atera fault line around the Dendahara district.

3 Microgravity survey

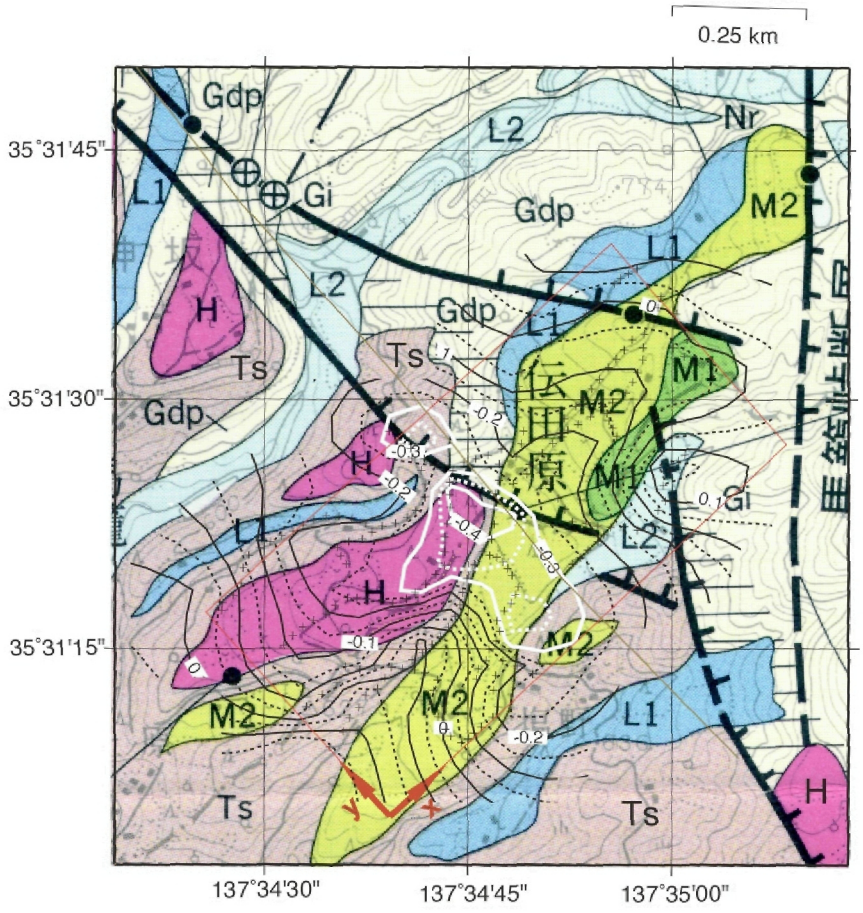


Fig. 2. Bouguer anomaly (Linear trend removed) on a strip map of the Atera fault (Tsukuda et al., 1993). Contour interval is 0.05 mGal. Symbol '+' means observation point. The Cartesian coordinate (arrows) used for Fig. 3. Abbreviations are as follows: H; higher terrace deposit, M; middle terrace deposit, L; lower terrace deposit, Ts; Toki Sand and Gravel, Gdp; Granodiorite porphyry, Gi; Inagawa Granite, Nr; Nohi Rhyolites. Tokyo datum is used as geodetic reference frame.

Measurements were carried out in the same manner as in the study of Tanaka et al. (2003) and the accuracy was better than 0.1 mGal through careful measurement, geodetic positioning of observation points, and reduction with high resolution elevation data. Figure 2 shows the Bouguer anomaly (linear trend subtracted) superimposed on a strip map of the Atera fault (Tsukuda et al., 1993). We adopted 2300 kg/m3 for the Bouguer density by determining the lowest

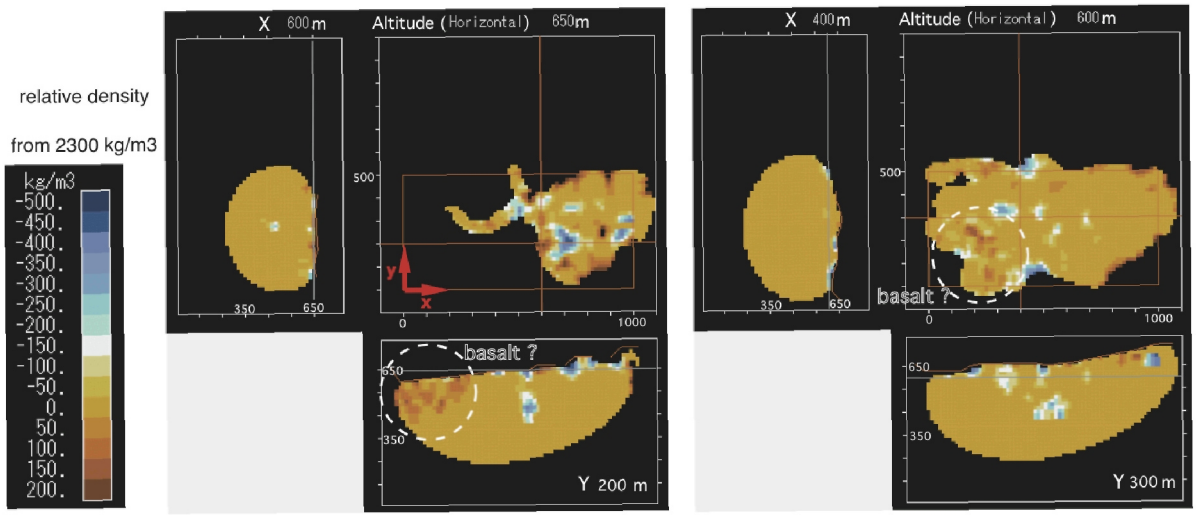


Fig. 3. Estimated density profiles by the method of Camacho et al. (2001, 2002). (Left) (x, y, z)=(600m, 200m, 650m), (right) (x, y, z)=(400m, 300m, 600m)

correlation with topography in the sediment distribution area. Accordingly, we investigated the structure under the sedimentary layer. Though basement rocks (granitic rocks and Nohi Rhyolites) seem to be near the surface under the sedimentary layer, density is thought to be less than 2500 kg/m3 due to fault induced fracturing, such as observed in the Kawaue borehole in the fracture zone of the Atera fault (Ikeda et al., 2001).

4 3-D gravity inversion

We performed a 3-D gravity inversion employing the technique of Camacho et al. (2001, 2002). This method can apply to an initial density model constructed by independent observation. A single density (2300 kg/m3) model was adopted, however, because no a priori information existed. Then, the density model is decided by the mixed condition on the residuals (fitness) and the whole anomalous mass (smoothness) with the variation of the density restricted to lie between 1800 and 2650 kg/m3. We assumed here that the lower and upper limit reflected terrace deposits and basaltic rock suffered from fracture, respectively. The density model that finally explained the observation value with a residual r.m.s. of 0.02 mGal is shown in. Fig. 3. There may be large errors near the edge of model. A low density

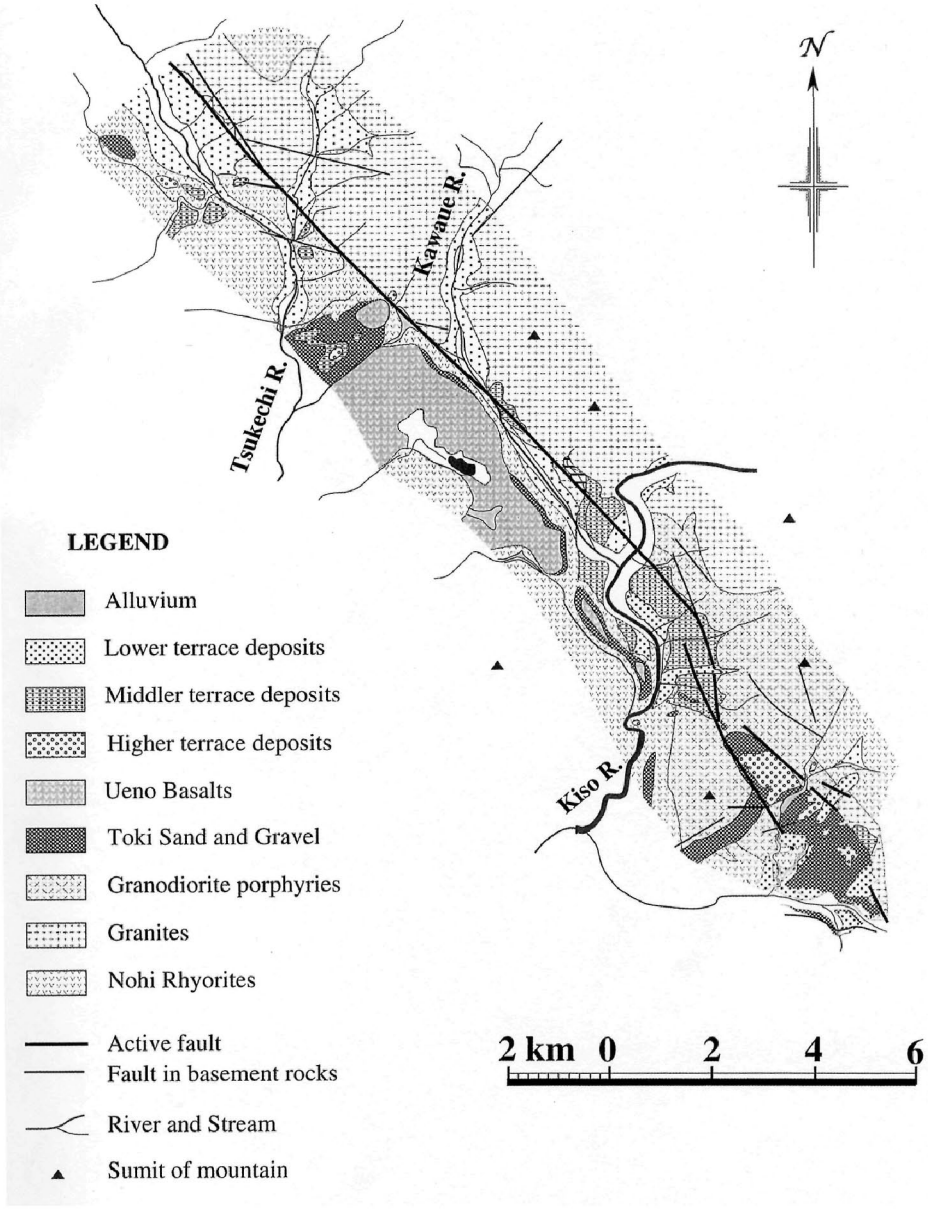


Fig. 4. Schematic map of the summary of this study (some additions from a geological map of the southern tip of the Atera fault by Yasue (2003)). The rectangular shows the microgravity survey area. Broken lines shows a crustal weak zone where basaltic magmas erupted, which was inferred by Tanaka et al. (2003) and its extrapolation to the SE direction.

zone corresponding to a fracture zone is distributed just under the trench site. However, a low density anomalous mass broadly distributed at shallow depth may be estimated as a lower density and smaller mass anomaly at greater depth due to the smoothness condition of Camacho et al. (2001, 2002). On the other hand, a high-density body seems to run about 200m southwest of the fault line (Fig. 3 right). Furthermore, the dip direction seems to slant toward the NE (Fig. 3 left). In any case, the shape of the dense body is consistent with it being associated with the Ueno volcano.

5 Conclusions

We carried out a microgravity survey in the Dendahara district with the aim of studying peculiar features along the Atera fault. As a result, the existence of a dense body (probably basaltic rock) was suggested to the southwest of the fault. We could not conclude whether this dense body is a part of the Ueno Basalts or not, but we presented the possibility that interaction occurred between dense rock and activity along the fault. Figure 4 summarizes the results of the study.

References

Camacho AG, Montesions FG, Vieira R (2002) A 3-D gravity inversion tool based on exploration of model possibilities. Computers and Geosci 28: 191-204

Camacho AG, Montesinos FG, Vieira R, Arnoso J (2001) Modelling of crustal anomalies of Lanzarote (Canary Islands) in light of gravity data. Geophys J Int 147: 403-414

Ikeda R, Omura K, Matsuda T, Mizuochi Y, Uehara D, Chiba A, Kikuchi A, Yamamoto T (2001) Central Japan's Atera active fault's wide-fractured zone: An examination of the structure and in-situ crustal stress. EOS, F1229, Nov. 20

Kimura J, Yoshida T (1999) Mantle diapir-induced arc volcanism: The Ueno basalts, Nomugi-toge and Hida volocanic suites, central Japan. The Island Arc 8: 304-322.

Nakata T, Imaizumi T, eds. (2002) Digital Active Fault Map of Japan, University of Tokyo Press (product serial number: DAFM1740) (in Japanese with English abstract)

Sugimura A, Matsuda T (1965) Atera fault and its displacement vectors. Geol Soc America Bull, 76: 509-522

Tanaka T, Aoki H, Martin AJ, Oshita K, Nozaki K, Onishi M (2003) Subsurface structure under a basaltic monogenetic volcano near the active Atera fault. Tectonophys (in press)

Toda S, Inoue D, Kubouchi A, Takase N, Nikaido M (1995) Paleoseismicity of the Atera fault system and 1586 Tensho Earthquake: Trenching studies at Ogo, Aonohara and Dendahara, Central Japan. Jour Seism Soc Japan, Second Series, 48: 401-421 (in Japanese with English abstract)

TRIES (2002) Geophysical survey of the Atera fault, central Japan. Report of Tono Research Institute of Earthquake Science 8, Association for the Development of Earthquake Prediction, Tokyo, Japan. (in Japanese)

Tsukuda E, Awata Y, Yamazaki H, Sugiyama Y, Shimokawa K, Mizuno K (1993) Strip map of the Atera fault system, scale 1:25000 and its explanatory text. Tectonic map series 7, Geological Survey Japan (in Japanese with English abstract)

Yasue K (2003) Developing process of active fault system in the eastern margin of inner zone of Southwest Japan: a case of the Atera Fault belt. PhD dissertation, Toyama University, Japan

Geothermal Reservoir Monitoring with Absolute Gravimetry

M. Sugihara
Geological Survey of Japan,
AIST, Tsukuba Central 7, 1-1-1, Higashi, Tsukuba, 305-8567, Japan

Abstract. A miniaturized absolute gravimeter, the FG5L was tested at Ogiri geothermal field. One day of measurements at a site provided a gravity value with an error of about 10 microGal. However, the fluctuations in the daily values greatly exceeded the daily estimated error. The FG5L is far from being useful for geothermal reservoir monitoring. The FG5, which was upgraded from the FG5L, was tested at Yanaizu-Nishiyama geothermal field. The FG5 requires only a few days to provide useful results at each survey location. Therefore one FG5 gravimeter can be used at all the geothermal fields in Japan that are being monitored. We have already used it at the 5 existing gravity networks.

Keywords. absolute gravimeter, FG5, FG5L, geothermal reservoir, gravity monitoring

1 Introduction

Microgravity monitoring is recognized as a valuable tool for mapping the redistribution of subsurface mass associated with geothermal exploitation. Generally, microgravity monitoring involves the measurement of small changes in gravity over time, across a network of stations, with respect to a fixed base. However, both local and regional gravity variations can cause errors in the determination of the gravity "datum" against which any measured changes are referred. A combination of absolute and relative gravimetry provides a solution to this problem.

Until recently, absolute gravimeters have been too cumbersome and expensive to use regularly at geothermal fields. However, a new miniaturized absolute gravimeter, Micro-g FG5L, was released onto the market in 2000. It is a simplified version of the standard absolute gravimeter FG5. The FG5L gravimeter offers an attractive trade-off between low price and precision in an instrument that is easy to use in the field. Many of the subcomponents of the FG5L are compatible with the FG5 in order to allow an upgrade path to the FG5. We introduced the FG5L(S/N 003) in January 2001, and then upgraded it to the FG5 (S/N 217) in January 2002. In the following, we report the performances of the two absolute gravimeters in Japanese geothermal fields.

2 Measurement by the FG5L

In February 2001, we carried out the first absolute gravity measurements with the FG5L/003 at the Ogiri geothermal field. A computer automatically controls the whole system with the standard software "g". The default settings are to start a series of measurements ("project") immediately and collect 48 sets of 100 drops per set every 30 minutes (at 15 seconds drop interval). On 12 February we carried out the first project. A project was performed every day until 17 February. This was followed by an uninterrupted 17-day run. The gravimeter, however, was found to have ceased functioning on the 14th day.

A comparison of actual recorded time with scheduled time, showed there had been a delay of 2 minutes, even on the first day (February 17), but it was recovered at the next set. On 22 February such delays began to happen in successive sets. However, the delay was shorter than five minutes so the next set could start at the scheduled time. Four days later (February 26) delays began occurring at intervals longer than five minutes, and could not be recovered. After February 28, some scheduled sets were skipped.

The lower part of the Figure 1 shows all the results including the uninterrupted project. During a one-day project or a 48-set the estimated value

of each set seems to be normally distributed and estimation error is around 10 microGal. However, the 48-set moving average that is almost identical to the interpolation of estimated value at each one-day project exhibits some trend and deviates by about 50 microGal from the mean. Looking at the variation of set values during the early four projects there are periodic patterns in the upper part of the Figure 1. The set values were corrected for solid earth tides, polar motion, and barometric pressure, but not for ocean tide loading. The Ogiri site is located 40 km northeast of the Sakurajima observatory, where the ocean loading effects were observed with the FG5 meter (Yamamoto et al., 2001). Referring it the periodic patterns in Figure 1 is too large to be attributed to ocean tide effect. The FG5L/003 was set on a concrete pier with a thickness of 1 meter at a location 300 m east of the power plant. A prefabricated house with an air conditioner was erected over this pier. We cannot draw confident conclusions regarding the periodic patterns; artificial seismic noise might cause them.

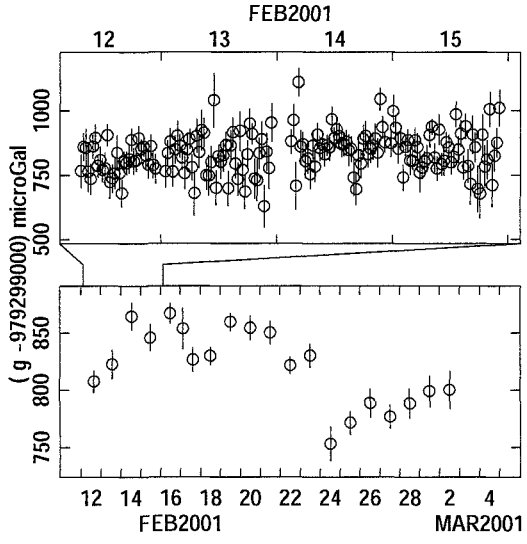


Fig. 1 Plot of absolute gravity measurements with the FG5L/003 at the Ogiri geothermal field. Set values of the early four projects (upper) and gravity changes during all the surveys (lower).

We chose to upgrade the instrument to FG5 specifications without carrying out more experiments. After upgrading we tested the FG5/217 with the small dropper which had been used with the FG5L/003 and observed anomalous gravity variation from day to day.

3 Measurement by the FG5

Since January 2002, the FG5/217 has performed for two purposes, geothermal reservoir monitoring and maintenance (Figure 2). Inter-comparison experiments have been made with other FG5 meters at sites of Mt. Tsukuba, Geographical Survey Institute (GSI) and whenever the instrument is returned to the manufacture in Colorado, USA. The measurements at Mt Tsukuba were made as a co-operative experiment among GSI, Earthquake Research Institute, Kyoto University and our group (Nishi et al., 2002). Before and after each geothermal field visit, measurements were made at the reference sites in GSI or in GSJ, in order to ensure that the absolute gravimeter gave consistent values.

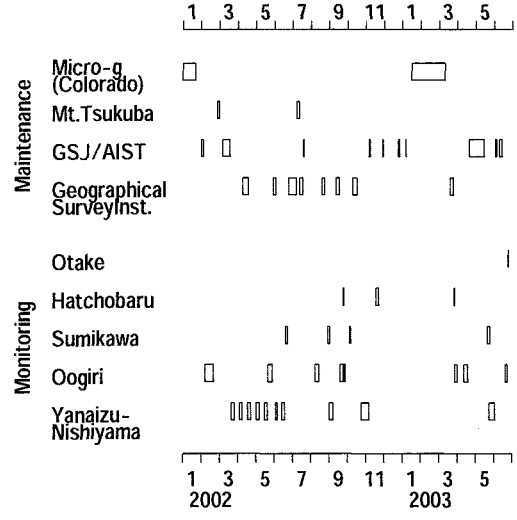


Fig. 2 History of the measurements with the FG5/217 absolute gravimeter.

In 2002 nine absolute gravity surveys were made with the FG5/217 at Yanaizu-Nishiyama geothermal field. A computer automatically controls the whole system with the software identical to the FG5L. The default settings are to start a series of measurements (a "project") immediately and collect 48 sets of 100 drops per set every 30 minutes (at 10 seconds drop interval). Each survey was made for at least 3 days, with the gravimeter being carefully set-up again at the start of each day ("project"), following the procedure shown by Williams et al. (2001). This not only produced significantly more data for a

given field measurement, but also allowed a comparison of the standard deviations of the half-hourly means ("set") each day with the variability from day to day. Figure 3 shows the relation among set, project and survey. The third survey consists of five projects and the 5th project consists of 29 sets. During the period of the 3rd survey a drilling pad for a new production well was being built near the station in the daytime, therefore absolute gravity measurements were made only during the night. At each survey an absolute gravity value was inferred from averaging all the set values.

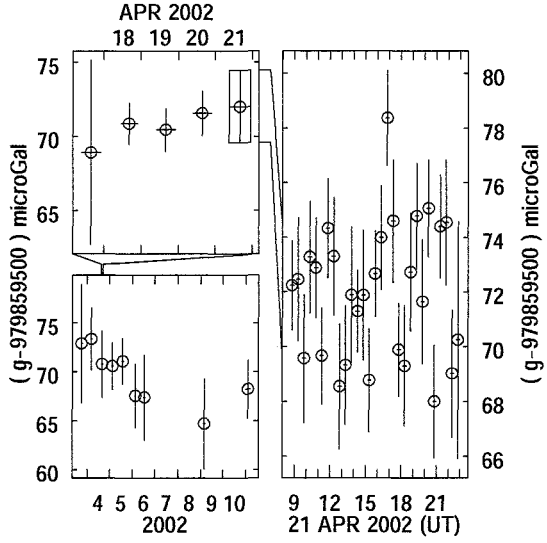


Fig. 3 Plot of absolute gravity measurements with the FG5/217 at the Yanaizu-Nishiyama geothermal field.

4 Relative and absolute measurement at Yanaizu-Nishiyama field

At the Yanaizu-Nishiyama field the New Energy and Industrial Technology Development Organization (NEDO) carried out gravity monitoring from 1997 to 2001. Their survey area was about 30 km^2 and covered the production zone and the reinjection zone with 138 stations. In 1998 short-term gravity changes were detected; these were related to a suspension in production (NEDO, 1999). In an effort to better understand the transient fluid flow associated with the suspension in production, we made a series of both the absolute and relative gravity measurements in and around the field in 2002. An initial survey was carried out before production was suspended.

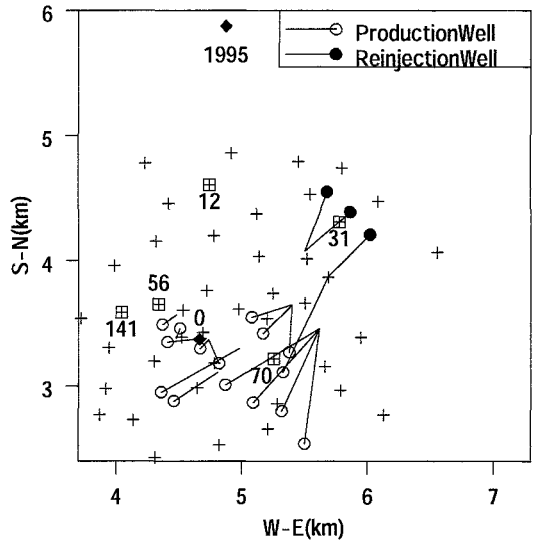


Fig. 4 Location of the gravity stations in the Yanaizu-Nishiyama geothermal field. The FG5/217 absolute gravimeter set at the point 0.

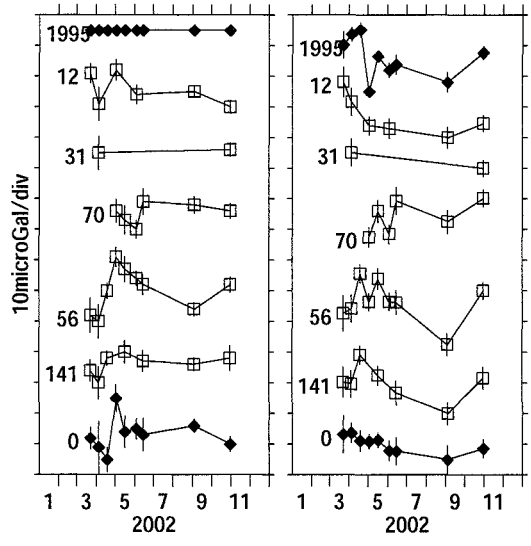


Fig. 5 Temporal gravity differences in 2002; to empirical datum (left), to absolute datum (right).

A Scintrex CG-3M gravimeter (S/N 385) was used to make relative gravity measurements at 58 points out of the NEDO's network (Figure 4), and the procedure was similar to that described by Ohta et al. (2001). The CG-3M gravimeter was operated in cycling mode at night alongside the

FG5 absolute gravimeter (at the point 0 in Figure 4). Each day's measurements were tied to the local base (#1995) situated outside the geothermal field. The gravity differences between the field stations and the local base (Figure 5a), and differences between the absolute station and the local base were determined. Data reduction techniques similar to those described by Ohta et al. (2001) have yielded preliminary gravity values with an estimated uncertainty +-5 microGal. By combining the absolute measurements with the relative measurements (Figure 5, left), the temporal gravity differences relative to an absolute reference (Figure 5, right) were determined.

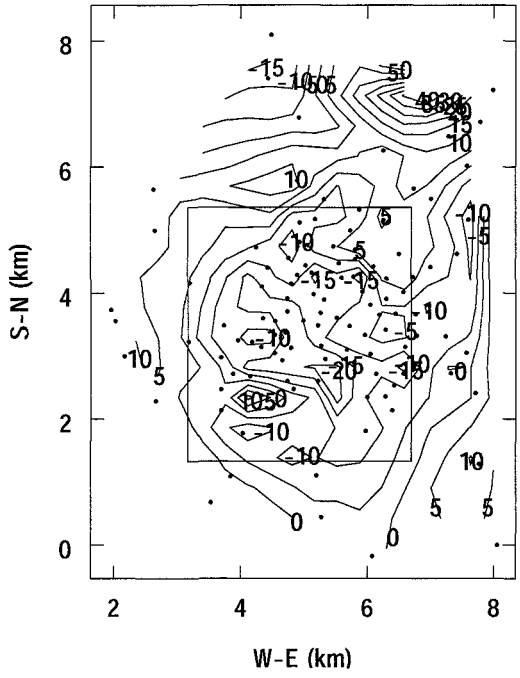


Fig. 6 Measured gravity change between 1997 and 2000 (modified from NEDO, 2002). The inset rectangle indicates the area of Figure 7.

In history-matching studies, observed gravity changes are compared with computed gravity changes arising from a particular mathematical reservoir model. Such comparisons will provide additional constraints on the history-matching process, and thus improve the quality and reliability of the reservoir model. The observed long-term gravity changes should constrain the representation of the reservoir boundary conditions (particularly the deep recharge boundaries) under conditions of fluid withdrawal. Figure 6 shows gravity changes observed in the field over the period between 1997 and 2000 by NEDO. Hashimoto et al. (2002) showed that the long-wavelength components of the observed gravity changes were reproduced by reservoir modeling (Figure 7). Long-term gravity changes at the absolute observation point is calculated to decrease with a rate of 6 microGal/Yr in 1998 and 4 microGal/Yr in 2002 by the latest model (NEDO, 2002). The observed gravity changes at the absolute observation point may be influenced by both long term changes and short term changes such as the effects of production suspension and the new production well. Short-term gravity changes have not been reproduced by reservoir modeling yet, however, the new observed gravity changes provide constraints on the more accurate history-matching process, and thus improve the reservoir model.

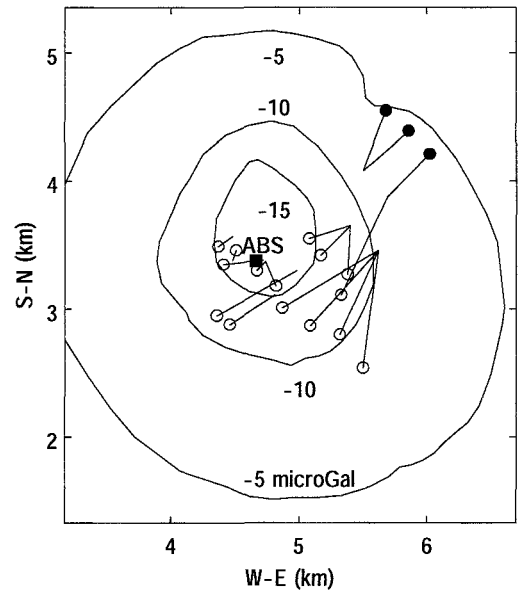


Fig. 7 Simulated gravity changes between 1997 and 2000 (modified from NEDO, 2002).

5 CONCLUSION

The goal of this program, introducing absolute gravimetry to geothermal reservoir monitoring, is to create absolute reference bases for the existing gravity networks in Japanese geothermal fields. Only a few days are required for an FG5 gravimeter to determine the gravity datum at each

survey at each network. Therefore one FG5 gravimeter can support observations at all the geothermal fields in Japan that are being monitored. We have already used at the 5 existing gravity networks (Figure 2).

Regarding the FG5L gravimeter it is not useful for geothermal reservoir monitoring. The accuracy of the FG5L was previously quoted as 20 microGals, but is now quoted as 50 microGals by the manufacturer. The repeatability is listed to be 10 microGals. The initial plan to operate the FG5L for a year for seasonal or monthly inter-comparison with a standard absolute gravimeter FG5. The results of the field test, however, show the repeatability of the FG5L/003 is far from being useful for geothermal reservoir monitoring.

References

Hashimoto, K., Horikoshi, T. And Kikuchi, T. (2002) Development of techniques for reservoir monitoring – current status of the project, *Proc. 24th New Zealand geothermal workshop*, pp.109-114.

NEDO (1999) Theme 2: Gravity monitoring, *The fiscal 1998 Report for the project of the development of technology for reservoir mass and heat flow characterization - monitoring and modeling of reservoir mass and heat flows.*

NEDO (2002) Theme 2: Gravity monitoring, *The fiscal 2001 Report for the project of the development of technology for reservoir mass and heat flow characterization - monitoring and modeling of reservoir mass and heat flows.*

Nishi, K., Hiraoka, Y., Kimura, I., Okubo, S., Furuya, M., Matsumoto, S., Fukuda, Y., Higashi, T. and Sugihara, M. (2002) Relative comparison of FG5 absolute gravimeter (2), *96th meeting abstracts*, the Geodetic Society of Japan.

Ohta, H., Yamanobe, H., Shiga, N., Tosha, T., Takemura, T. and Kobayashi, E. (2001) Gravity monitoring of the Yanaizu-Nishiyama geothermal field, Japan, *GRC Trans.*, 25, pp.669-674.

Williams, S. D. P., Baker, T. F. and Jeffries, G. (2001) Absolute gravity measurements at UK tide gauges, *Geophys. Res. Lett.*, 28, pp.2317-2320.

Yamamoto, K., Ishihara, K., Okubo, S. and Araya, A. (2001) Accurate evaluation of ocean tide loading effects for gravity in nearshore region: the FG5 measurements at Sakurajima volcano in Kagoshima Bay, Japan, *Geophys. Res. Lett.*, 28, pp. 1807-1810.

Towards a rigorous combination of VLBI and GPS using the CONT02 campaign

D. THALLER, R. SCHMID, M. ROTHACHER
Forschungseinrichtung Satellitengeodäsie (FESG), TU München, Arcisstr. 21, 80290 München, Germany

V. TESMER, D. ANGERMANN
Deutsches Geodätisches Forschungsinstitut (DGFI), Marstallplatz 8, 80539 München, Germany

Email: daniela.thaller@bv.tum.de
Phone: ++49-89-289-23186
Fax: ++49-89-289-23178

Abstract. The International VLBI Service for Geodesy and Astrometry (IVS) set up a 15-day campaign of continuous VLBI observations, named CONT02, that took place October 16-31, 2002. The goal of CONT02 was to acquire state of the art VLBI data over a continuous two-week period to demonstrate the highest accuracy of which VLBI is capable. Using the two weeks of CONT02 data from VLBI and the global GPS data for the same time span, daily unconstrained normal equations were generated taking much care to use identical models and the same parameterization of the common parameters for both space geodetic techniques. Based on these normal equations first steps towards a fully rigorous combination of all common parameters were performed by including station coordinates and Earth orientation parameters (EOP) into the solutions. In addition, the impact of the local tie information on the combined solutions was studied. To assess the quality of the individual and combined solutions, polar motion and UT1-UTC parameters were set up with a two-hour time resolution and the results were compared with sub-daily models for ocean tide variations derived from altimetry and other space geodetic techniques.

Keywords: *CONT02, Sub-daily EOPs, Local ties, Troposphere*

1 Introduction

The two-weeks VLBI campaign named CONT02 that was initiated by the IVS took place October 16-31, 2002. Altogether, eight stations participated in this campaign, namely Algonquin, Fairbanks, Hartebeesthoek, Kokee Park, Ny-Alesund, Onsala, Westford and Wettzell.

Using the two weeks of CONT02 data from VLBI and the GPS data from about 150 globally distributed IGS stations for the same time span, daily unconstrained normal equations were generated for both techniques taking much care to use identical models (e.g. solid Earth tides, ocean loading, a priori hydrostatic delay, mapping function for the zenith delay, a priori nutation model, a priori interpolated values of EOPs,...) and the same parameterization of the common estimated parameters:

- one set of station coordinates per solution, i.e., for each day,
- two-hour pole coordinates and UT1-UTC,
- daily nutation parameters,
- two-hour troposphere zenith delays per station (as constant offsets),
- daily troposphere gradients per station.

The analysis of the VLBI observations was done with OCCAM (Titov et al. 2001), and the Bernese GPS Software (Hugentobler et al. 2001) was used for the GPS analysis. More details concerning the GPS analysis can be found in the summary of the CODE analysis strategy located in the ftp-archive of the IGS Central Bureau at

igscb.jpl.nasa.gov/igscb/center/analysis/code.acn.

The adaptation to common models and parameterizations for the analysis of the observations was possible due to the very close cooperation between the two participating institutions and can be seen as a big advantage for these combination studies because normally each analysis center uses its own strategy. The solution series available so far for combination tests (e.g. in the data pool for the IERS SINEX Combination Campaign, see Rothacher et al. 2003, Angermann et al. 2003) are based on different assumptions and therefore the interpretation of the comparison and combination results is more complicated.

For the CONT02 campaign both, an analysis of the single-technique solutions as well as a daily inter-technique combination were performed. In a first step

the common Earth orientation parameters (EOP) and the local tie information at the eight co-located stations were used to connect the two techniques and the impact of the local tie information on the combination results was studied.

2 Comparison of VLBI and GPS solution series

Before the inter-technique combination was done, each technique solution was analysed separately and the common parameters were compared.

There are only a few criteria for the validation of the results and the evaluation whether a combination is better than the single technique solutions: the repeatability of station coordinates, a comparison with subdaily ERP models derived from independent techniques, and the data of water vapour radiometers for the troposphere parameters.

2.1 Repeatability of station coordinates

For the VLBI network a repeatability of 6.3 mm, 4.3 mm and 6.9 mm for the north, east and height component, respectively, can be achieved over the two weeks if all data are used. The relatively bad repeatability for the north component results from a bad performance of the Wettzell observations for a few days.

The analysis of the global GPS network shows a repeatability for the station coordinates of 2.1 mm, 2.1 mm and 6.3 mm for the north, east and height component, respectively.

It can be seen that the height component is the most critical for both techniques, mainly because of tropospheric influences. The quality in the height determination is comparable for VLBI and GPS whereas the horizontal components are better determined in the GPS solution. It should be kept in mind, however, that the GPS solution contains about 150 stations whereas only eight stations contributed to the VLBI network.

2.2 Earth rotation parameters

In the sub-daily estimates of the Earth rotation parameters (ERP) some remaining periodical differences compared to the C04 series are visible (see Fig. 1). These variations are in good agreement between the two techniques. Only the absolute values for the y-pole are shifted by about 0.4 mas for the VLBI solution series compared to the GPS solution series. This bias may result from a reference frame inconsistency between GPS and VLBI together with the sparse global coverage of the VLBI network (no stations in Asia and Australia, only one station in the Southern hemisphere, altogether only eight stations). In addition, it can be seen that the scattering in the VLBI results is larger than in the GPS results. How critical the network configuration in the VLBI solution is becomes more evident if the period around day 300 is considered: the station Algonquin was missing between day 299, 18:00 UTC and day 300, 18:00 UTC, and at least in UT1-UTC and in the x-pole some clear peaks can be recognized for this time span.

2.3 Troposphere parameters

The troposphere parameters, zenith delays as well as gradients, were included in both technique solutions. The dry part of the zenith delay was modelled according to a standard atmosphere, the Saastamoinen model and the Niell dry mapping function, and the remaining wet part was estimated using the Niell wet mapping function. As the dry part was always modelled with respect to the GPS reference point, the estimated zenith delay for the VLBI solution does not only include the wet part but also a very small correction for the dry part stemming from the height difference between the GPS and VLBI reference points. As an example, Fig. 2 shows the two-hour troposphere zenith delays w.r.t. the GPS reference point for the station Algonquin estimated independently from VLBI and GPS as correction to the a priori dry part. Even the small sub-daily variations are recognized nearly identical in the results of both techniques.

According to a simple theoretical consideration for small height differences, a difference of 10 m in station height corresponds to about 3 mm difference in the hydrostatic part of the troposphere zenith delay. Comparing these theoretical values with the biases derived from the VLBI and GPS troposphere estimates (Table 1) it can be seen that not all stations are in good agreement with this theoretical bias. The mean estimated difference is always larger than the theoretical bias which means that either the zenith delays derived from GPS are systematically too large or the VLBI results are too small. At the moment, a correlation of the discrepancies with the height difference, the size of the VLBI antenna or the GPS antenna type cannot be recognized. A possible explanation for a common systematic bias might be the use of relative antenna phase center variations for GPS satellites and antennas whereas absolute phase center variations would decrease the troposphere zenith delays by about 4 to 5 mm (Schmid et al. 2003). In addition, the question which point should be considered as reference for the troposphere estimation in VLBI is not yet answered. Up to now, the intersection of the axes is usually taken as

reference point for the a priori dry delay but this is certainly not correct. The point the VLBI troposphere estimates are referring to is virtually realized by the separation of the parameters "station height", "station clock offset" and "troposphere zenith delay" due to the partial derivatives w.r.t. these parameters in the analysis software. Therefore, common delays to all quasar observations are interpreted as "clock" whereas the part of the delay that varies with *1/cos(z)* is considered to be part of the tropospheric delay.
Table 1 shows as well the rms of the differences between the troposphere estimates derived separately from VLBI and GPS.

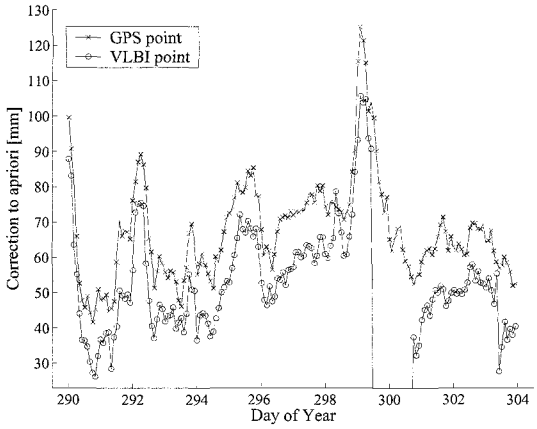


Fig. 2: Two-hour troposphere zenith delays for the station Algonquin, estimated independently from VLBI and GPS as correction to the a priori dry part for the GPS reference point. For the time span between day 299, 18:00 UTC and day 300, 18:00 UTC the station Algonquin is missing in the VLBI analysis.

Table 1: Differences in the estimated troposphere zenith delays between VLBI and GPS (as mean values over two weeks) compared to the theoretical troposphere bias derived from the height differences between the reference points and the rms of these troposphere differences.

Site	Mean difference [mm]	Theory [mm]	Height difference [m]	RMS [mm]
ALGO	14.0	6.9	23.113	4.4
FAIR	6.8	3.9	13.082	5.9
HRAO	7.8	0.5	1.543	6.9
KOKB	14.2	2.8	9.236	8.4
NYA1	1.7	0.9	3.071	2.5
ONSA	5.1	4.1	13.713	4.6
WES2	11.7	0.5	1.753	5.0
WTZR	4.5	0.9	3.097	6.0

The troposphere gradients for the north and east direction were estimated once per day for each station and the results for the co-located stations are in good agreement between the two techniques. As an example, Fig. 3 shows the results for the station Onsala.

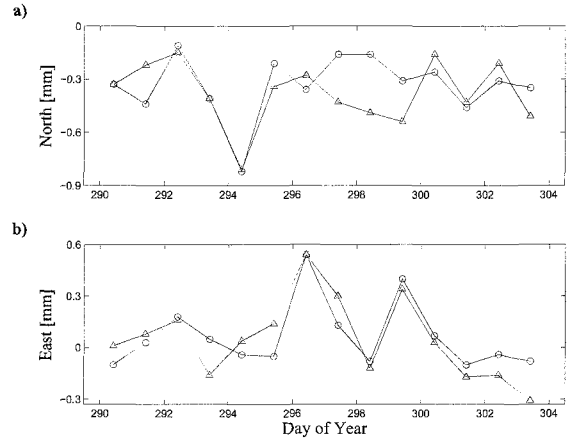


Fig. 3: Daily constant troposphere gradients for Onsala derived from GPS (= triangles) and VLBI (= circles): **a)** North direction, **b)** East direction.

3 Inter-technique combination

The connection between VLBI and GPS for the daily inter-technique combination was realized on one hand by the common EOPs and on the other hand by the local ties for the co-located sites (for Ny-Alesund see Steinforth et al. 2003; else, the ITRF2000 local ties were used, Altamimi et al. 2002) that were introduced as pseudo-observations with appropriate weights. For the initial results presented here, the troposphere parameters were not yet combined.

The normal equations of VLBI and GPS were weighted according to the station coordinate repeatability of each technique, with one common weighting factor for all daily normal equations of the two weeks.

3.1 Selection of local ties

A critical step in the inter-technique combination is the selection of good local ties. Different daily combinations were analysed for this purpose:

- introducing all local ties with the same weight,
- excluding one or more local ties that seem to be questionable,
- using different weights to cope with the quality of the local tie information.

Several criteria were applied to decide which local tie information seems to be suitable: the resulting coordinate differences compared to the a priori local tie information, the repeatability of the station coordinates, and the results for the EOPs.

As outcome of all these investigations, the local tie for Kokee Park seems to be not suitable for the combination and some other local ties (Fairbanks, Hartebeesthoek, Algonquin) do not fit very well to the VLBI and GPS results for CONT02.

It is worth mentioning how big the influence of only one bad local tie can be, e.g., on the ERPs: Fig. 4 shows the results for the x-pole for different combinations, one including all local ties and others with some local ties excluded. It can be seen that the differences between these solutions are not negligible at all and that especially the local tie for Kokee Park has a large influence on the results for the ERPs. The results for the solutions with resp. without the local tie for Kokee Park are shifted by about 0.2 mas. Compared to this effect, the impact of the other local ties that might not be very good either is marginal and the ERP results do not differ very much. During the analysis it was also recognized that the number of co-located sites used for connecting the different techniques is very critical and eight stations (which is the maximum for CONT02) are on the lower limit to define a stable common reference frame, and dropping one local tie must be considered very well.

For all the following combination studies the local tie for Kokee Park was excluded and the seven other local ties were introduced in the combination with equal weights.

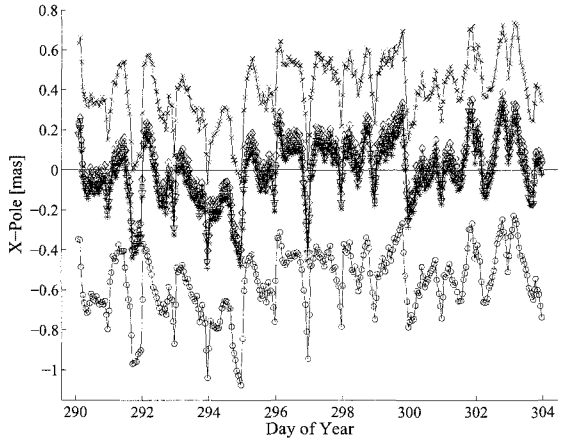


Fig. 4: Results for the x-pole derived from combinations of VLBI and GPS using different sets of local ties, compared to C04: all local ties used (crosses); w/o KOKB (squares); w/o KOKB+HRAO (stars); w/o KOKB+FAIR (diamonds); w/o KOKB+FAIR+HRAO (triangles); w/o KOKB and FAIR+HRAO down-weighted (circles).

3.2 Combination results

As a primary result concerning the station coordinates it turned out that the inter-technique combination improved the repeatability for the co-located sites with good local tie information, especially for the VLBI stations because due to the combination they are embedded in the large and more stable GPS network. Table 2 summarizes the repeatabilities for the VLBI and GPS stations with good local ties derived from the single technique solutions and the combination. The large improvement in the height component of the VLBI stations is astonishing and the reason is not yet completely clear. Therefore further investigations are needed in this field and these results have to be proved by other groups and other combination studies, e.g. the IERS SINEX Combination Campaign (Rothacher et al. 2003).

Table 2: Repeatability for co-located stations with good local ties derived from the single technique solutions and from the combination [mm].

Solution type	North	East	Height
VLBI only	6.8	4.4	7.1
VLBI from combination	4.2	3.0	3.5
GPS only	1.4	1.6	5.6
GPS from combination	2.5	2.4	5.3

The combined ERPs are shown in Fig. 5 and the remaining daily and sub-daily variations compared to the C04 series that were already visible in the single solution series can be seen in the combination as well. These periodic differences w.r.t. C04 might be caused by missing terms in the sub-daily ERP model. The Ray96 model (see McCarthy 1996) was used in the computations but it only includes the eight principle terms. Another sub-daily model, developed at FESG and very similar to the model of the IERS Conventions 2000, was derived from about 20 years of VLBI data and 7 years of GPS data and includes about 50 terms. The differences between the two models are shown in Fig. 5 as well. We hoped to see a better agreement between the remaining sub-daily variations of the combined compared to the individual solutions with the model improvements from Ray96 to the more involved sub-daily model, but this is not evident. The amplitudes of the residual variations in the combination results are always larger than the differences between the two models. One problem might be the difficulty to separate sub-daily ERP variations and the daily nutation estimates. Besides these sub-daily variations, the results for the

y-pole are shifted by about 0.161 mas w.r.t. C04. This inconsistency was already recognized in the IERS EOP Alignment Campaign (Rothacher et al. 2003) and the IERS SINEX Combination Campaign (Rothacher et al. 2003) so that the results from the CONT02 analysis are in good agreement with the outcome of these campaigns.

4 Conclusions

As a small step towards the concept of IGGOS the CONT02 campaign was used as a test case for the rigorous combination of common parameters in VLBI and GPS on the normal equation level. The analysis of the series of common parameters has shown that not only station coordinates and EOPs but also the troposphere parameters can be combined. The best way for an inter-technique troposphere combination still has to be studied and the impact of such an extended combination on all common parameters should be investigated next.

Concerning the local tie information two results must be pointed out: first, the selection of the local ties that are used in the combination must be done very carefully, especially if there are only a few co-located sites available like for the CONT02 campaign, because the impact of only one bad local tie on the combined solution is enormous, as seen for the pole coordinates. Second, it seems that good local tie information can improve the repeatability for these co-located station coordinates and especially the height component benefits from this additional information. But, as already mentioned, these results still have to be confirmed in other studies. In addition it must be mentioned, that a comparison with results for the local ties derived from other computations, e.g. ITRF2000 (Altamimi et al. 2002), is not as simple because the ITRF is derived from data over a long time span and then a local tie might look worse (e.g. due to antenna changes for GPS) whereas it might fit quite well to the observations if only two weeks are analysed.

The two-hour resolution for the ERPs indicates that there are some periodic differences compared to C04 but a comparison of sub-daily ERP models did not prove that these remaining variations can be explained solely with deficiencies in the used sub-daily model, but might also result from the difficulty to separate nutation offsets and sub-daily ERPs.

Acknowledgements. The funding for the most part of this work is given by the German Bundesministerium für Bildung und Forschung (BMBF) within the Geotechnologien project. We like to thank as well the group of the Geodetic Institute of the University of Bonn (GIUB) for making available the VLBI - GPS local tie for the station Ny-Alesund, and the group of the CODE analysis center at the Astronomical Institute of the University of Berne (AIUB) for providing the pre-processed GPS data.

References

Altamimi Z., P. Sillard, C. Boucher (2002) ITRF2000: A new Release of the International Terrestrial Reference Frame for Earth Science Applications. J. Geophys. Res., 107 (B19), 2214, doi : 10.1029/2001JB000561.

Angermann D., D. Thaller, M. Rothacher (2003) Position Paper: IERS SINEX Combination Campaign, IERS Technical Note 30 (in press), Bundesamt für Kartographie und Geodäsie, Frankfurt am Main, Germany.

Hugentobler U., S. Schaer, P. Fridez (eds.) (2001) Bernese GPS Software Version 4.2, Astronomical Institute, University of Berne, Switzerland.

McCarthy D. (ed.) (1996) IERS Conventions, IERS Technical Note 21, Observatoire de Paris, Paris, France.

Rothacher M., R. Dill, D. Thaller (2003) IERS Analysis Coordination, in: Dick W., B. Richter (eds.) IERS Annual Report 2002, Bundesamt für Kartographie und Geodäsie, Frankfurt am Main, Germany.

Schmid R., M. Rothacher (2003) Estimation of elevation-dependent satellite antenna phase center variations of GPS satellites, Journal of Geodesy, accepted for publication.

Steinforth C., R. Haas, M. Lidberg, A. Nothnagel (2003) Stability of Space Geodetic Reference Points at Ny-Alesund and their Excentricity Vectors. To be published in: Thorandt V. et al., Proceedings of the 16th Working Meeting on European VLBI for Geodesy and Astrometry, Bundesamt für Kartographie und Geodäsie, Leipzig, Germany.

Titov O., V. Tesmer, J. Böhm (2001) OCCAM 5.0 users guide, AUSLIG Technical Report 7, AUSLIG, Canberra, Australia.

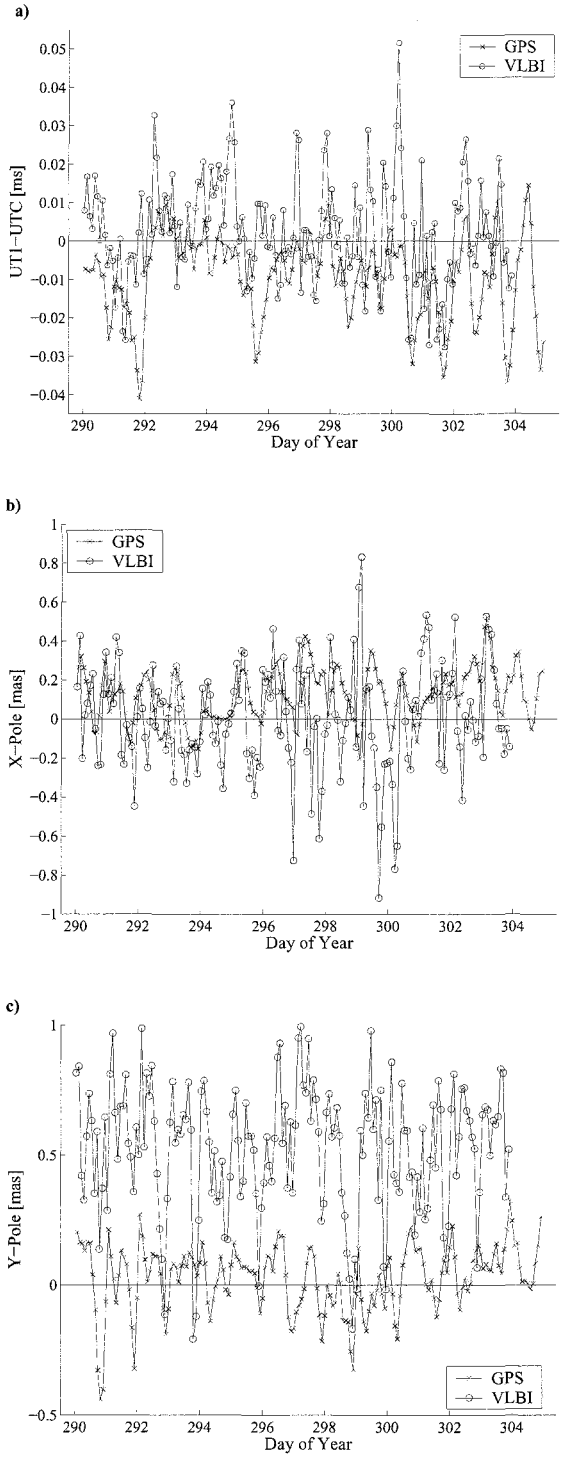


Fig. 1: Two-hour ERPs derived from VLBI and GPS compared to the C04 series:
a) UT1-UTC, **b)** X-Pole, **c)** Y-Pole.

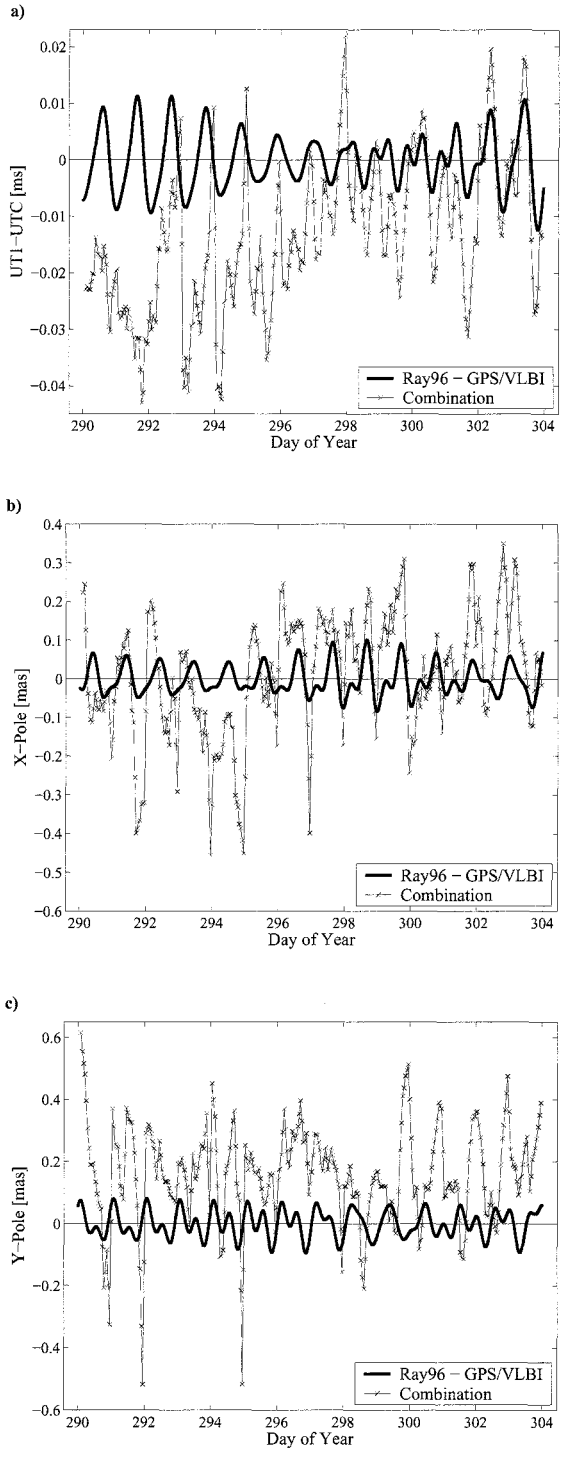


Fig. 5: Combined ERPs compared to C04 and differences between the sub-daily ERP models derived by Ray 1996 and at FESG:
a) UT1-UTC, **b)** X-Pole, **c)** Y-Pole.

Fuzzy Inference Systems for the Prediction of Earth Rotation Parameters

Orhan Akyilmaz[1], Hansjörg Kutterer[2]
[1] Istanbul Technical University, Division of Geodesy, 34469 Maslak, Istanbul, Turkey
[2] Deutsches Geodätisches Forschungsinstitut, Marstallplatz 8, D-80539 Munich, Germany
E-Mail: akyilma2@itu.edu.tr

Abstract. The paper presents short-term and long-term predictions of Earth rotation parameters (ERP) (Length-of-Day and polar motion) up to one year by means of ANFIS (Adaptive Network based Fuzzy Inference System). The ERP C04 time series of the International Earth Rotation and Reference Systems Service (IERS) with daily values are used as data basis. The impact of the tides of the solid Earth and the oceans and seasonal variations of the atmosphere are removed from the C04 series. The residual series are used both for training and validation of the ANFIS. For network optimization purposes, different architectures are considered. The results of the prediction are compared with those from other methods. The ERP values predicted by ANFIS have RMS errors, which are equal or even less than those from other methods. The most striking advantage of using ANFIS instead of other approaches is its significantly reduced complexity.

Keywords. Prediction, Earth Rotation, Fuzzy Inference Systems, Soft Computing, ANFIS

1 Introduction

Today, daily or even subdaily time series of Earth rotation parameters (ERP) are available with high precision from the International Earth Rotation and Reference Systems Service (IERS). ERP series are important for many investigations and applications in geodesy and astronomy since they link celestial and terrestrial reference frames. As these parameters are determined by space-geodetic techniques such as Global Positioning System (GPS), Very Long Baseline Interferometry (VLBI), Satellite Laser Ranging (SLR), etc., they are not available in real-time. For navigation purposes it is necessary to predict the ERP at least over a few days.

Up to now, various prediction methods have been developed, e.g. by Zhu (1981, 1982), Chao (1984), McCarthy and Luzum (1991), Freedman et al. (1994), McCarthy (1996), Malkin and Skurikhina (1996), Ulrich (2000), Schuh et al. (2002). In this study, a previous presentation (Akyilmaz and Kutterer, 2003) is extended regarding the long-term prediction. The paper concludes with a comparison of the new results and the ones discussed by Schuh et al. (2002).

The Adaptive Network based Fuzzy Inference System (ANFIS) which is used here will be explained in Section 2. It is primarily based on fuzzy set theory, which was introduced by Zadeh (1965). Fuzzy set theory and the related fuzzy logic are widely applied in various disciplines. Their main idea is the extension of crisp or sharp boundaries of classical sets as well as of logical variables and expressions so that any element has a continuous degree of membership between [0, 1] instead of the binary range {0,1} in classical set theory.

For a typical membership function (MF), i.e. the generalized bell function, see Fig. 1. The functional description is given in Eq. (1).

$$m(x) = \frac{1}{1+\left[\left(\frac{x-c}{a}\right)^2\right]^b} \tag{1}$$

The values of x with m(x)=1 are full elements of the fuzzy set, those with m(x)=0 are not element and all others are element just to a certain degree.

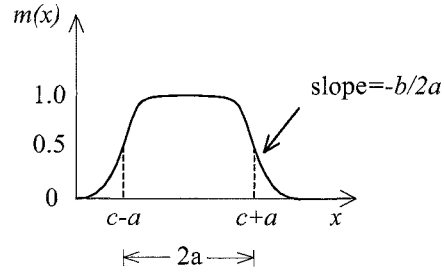


Fig. 1. Meanings of the parameters in the generalized bell membership function

2 Fuzzy Inference Systems

Fuzzy inference systems transform fuzzy information from an input space ('premises') to an output space ('consequents') by means of fuzzy if-then rules (Akyilmaz and Kutterer 2003). Fuzzy if-then rules are conditional statements of the form IF **A** THEN **B**, where **A** and **B** are labels of fuzzy sets characterized by appropriate membership functions (Zadeh 1965). They may differ only in the types of membership functions in the consequent parts. Here the type proposed by Takagi and Sugeno (1983) is used which is defined by a non-fuzzy equation of the input variables.

Adaptive Network based Fuzzy Inference Systems (ANFIS) have been developed by Jang (1993). ANFIS are functionally equivalent to fuzzy inference systems. They use a supervised learning algorithm to optimize the parameters of an inference system based on fuzzy logic. The basic idea can be described as follows: a fuzzy inference system is typically defined by the membership functions of linguistic input and output variables and an inference rule base. As such a system is just an initial guess for an adequate model, the system performance has to be optimized by tuning the premise and consequent parameters using the actual data.

In this paper, the First-Order Takagi-Sugeno ANFIS is used in which the consequent part is a linear function of the input variables. The supervised learning strategy is based on a hybrid algorithm, which is a combination of gradient descent and Kalman filter. The advantage of this update algorithm is the very fast convergence and the guarantee for reaching the global optimum. For more details on the algorithmic aspects of ANFIS see (Akyilmaz and Kutterer 2003).

3 Prediction of ERP by ANFIS

In order to make a prediction of ERP by ANFIS, the data have to be arranged as input-output pairs. Here, the time series itself has been used for composing both the input matrices and the output vectors. The output of each ANFIS model for the prediction of an individual day in future is a vector since particular models have been developed to predict the respective ERP value such as the one of the 1st day, 2nd day and so on. This has been employed in order to reduce the training time and to improve the results by eliminating some inner constraints of the fuzzy inference system. The model parameters are optimized using the training set. They are confirmed by means of the validation set.

3.1 Reduction of the time series and generation of training patterns

Daily values of the IERS C04 series between 1980 and 2001 were used for training and validating of the composed fuzzy models. For the short-term predictions, the entire data set was divided into two parts: the values from 1980 to 1997 were used for training and the remaining values from 1998 to 2001 for validating the network.

Since a fuzzy inference system requires the validation set to be covered by the range of the training set, a linear trend was reduced from the polar motion series. Hence, the errors which may come from extrapolating are eliminated to some extent (see Fig. 2). To this end the parameters of a linear best-fitting function $[\,\mathrm{model}(t) = a_0 + a_1 t\,]$ were estimated by the least-squares method individually for the x and y components of polar motion from the time series between 1980 and 1997. Then the functional values of these linear functions were extrapolated until 2001. Furtheron, the linear functions were subtracted from the actual time series for the whole interval between 1980 and 2001. The remaining data have been used in training patterns.

In case of LOD, tidal effects were removed according to the IERS Conventions, Chapter 8 (McCarthy 1996). Similar to the procedure employed for polar motion components, a linear trend was reduced by means of the least-squares method and extrapolated until 2001. The residual series were used for training and validating of the network. ERP before and after the mentioned reductions are given in Fig. 2.

For the long-term predictions the IERS C04 data between 1978 and 1999 have been taken and the past 80% were used for training whereas the remaining 20% were used to confirm the composed fuzzy models. This has been done in order to make a consistent comparison with Schuh et al. (2002). Individual fuzzy models are composed for the next 6, 9 and 12 months in future as examples for long-term predictions.

In contrast to other works (Ulrich 2000, Schuh et al. 2002) no further signal components were reduced a priori.

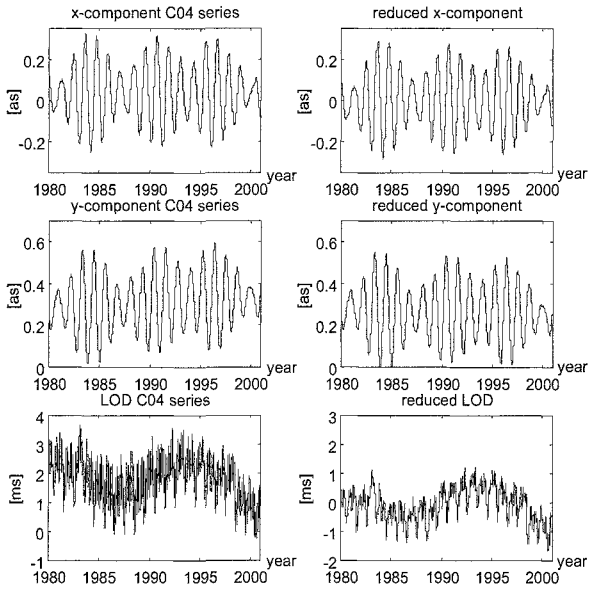


Fig. 2. ERP before and after reduction process

After the ERP components had been reduced, the following training patterns were generated:

Patterns for polar motion components x and y:

Let x(t) be the value of the time series element to be predicted. This means that x(t) is the element of the output vector. Additionally, let k be the number indicating the day in future to be predicted (k = 1, 2,...). For the short-term predictions the considered pattern is composed as

$$\{x(t-8k), x(t-4k), x(t-2k), x(t-k)\} \rightarrow \{x(t)\}.$$

For long-term predictions, the relevant patterns are generated simply in the following form

$$\{x(t-4k), x(t-3k), x(t-2k), x(t-k)\} \rightarrow \{x(t)\}.$$

These patterns are then shifted along the whole time series of both reduced polar motion components. Obviously, the training patterns for polar motion components consist in both cases of four inputs and one output for each prediction day.

Patterns for LOD:

Since the variation of LOD is much more irregular and complicated than those of the polar motion components, patterns were composed using five inputs for short-term prediction. The values of the reduced LOD time series of the last five days were selected as inputs and the day to be predicted was selected as output. Again a particular pattern was composed for each day to be predicted. Some of these patterns for the prediction of the first ten days in future are given below.

$$\{LOD(t-5), ..., LOD(t-1)\} \rightarrow \{LOD(t)\} \quad k = 1$$
$$\{LOD(t-6), ..., LOD(t-2)\} \rightarrow \{LOD(t)\} \quad k = 2$$
$$\vdots$$
$$\{LOD(t-14), ..., LOD(t-10)\} \rightarrow \{LOD(t)\} \quad k = 10$$

However, for the long-term LOD prediction a pattern with just four inputs was used. Though a higher number of inputs would have increased the accuracy of prediction, it was not considered in this study in favour of a high efficiency of the fuzzy inference systems. Here, "high efficiency" denotes a low complexity and a short computer run time and at the same time a reasonable accuracy of the prediction.

The patterns for long-term predictions were generated as

$$\{LOD(t-4k), ..., LOD(t-k)\} \rightarrow \{LOD(t)\}$$

with k=180, 270, 360. These pattern matrices are then shifted along the whole time series of the reduced LOD forming a multitude of pattern pairs.

3.2 Description of the fuzzy sets and network training

After composing the training patterns, each input variable space has to be covered with fuzzy sets (w.r.t. membership functions) each of which is represented by a linguistic label like, e.g., "small", "large", etc. In fact, the optimum network configuration depends on both the composed patterns and the number and the type of the membership functions. One has to keep these two factors in mind while seeking the best configuration.

There are some restrictions such as calculation time and overfitting phenomena, which bound the meaningful number of membership functions. The higher the number of membership functions the more parameters are to be estimated what may increase the calculation time significantly. A network that is too complex may fit the noise, not just the signal, leading to overfitting. Overfitting is especially dangerous because it can easily lead to meaningless predictions that are far beyond the range of the training data.

In practice, one tries to compose an ANFIS network with as few parameters as possible what

leads to the smallest possible number of fuzzy if-then rules. Additionally, the number of fuzzy if-then rules is directly related to the number of membership functions in each variable space. For example, if there are four variables in input space and each one is represented by two membership functions, then the number of fuzzy if-then rules is equal to 2^4=16. An extra membership function in each variable space will lead to 3^4=81 number of fuzzy rules.

Here, for short-term predictions, each variable of the input space is represented by two membership functions of Gaussian type (Akyilmaz and Kutterer 2003). Thus, the number of fuzzy if-then rules in the ANFIS networks established for short-term polar motion predictions is 16 whereas it is 32 for short-term LOD predictions. Note that the predicted residual value of the fuzzy inference system for any particular day has to be added to the corresponding value of the linear trend function to obtain the final value to be predicted.

For accurate long-term predictions of polar motion, a higher number of membership functions is required. This is because the distances between input vectors are too far and it is not adequate to describe the relationship between inputs and outputs by using two MF. Thus, three MF for each input variable is considered. For long-term LOD predictions, the number of two-input MF is still kept. However, the number of rules is the half of that for short-term predictions, i.e., 2^4=16.

The hybrid learning algorithm, which was briefly mentioned in Section 2, has been used for training each network. The output function is selected as a constant variable which is called singleton instead of a linear function of input variables in order to reduce the calculation time and the errors in case of using predicted values as inputs for the further predictions. Singletons are fixed-value numbers after training.

Actually, there were no significant differences between using constant or linear output functions. For each ANFIS prediction model less than 40 iterations were needed for the convergence to the global minimum value of the sum of the squared errors for the network configuration in question. This corresponds with a calculation time of a few minutes in a 64 RAM Pentium II-350 CPU computer. This reduction is one of the most important advantages of the ANFIS modeling.

A typical example of a training run is plotted in Fig. 3. As expected, the error of the training set (black) is decreasing with increasing number of training cycles. One interesting point in this figure is that after reaching the global minimum in the training cycles, the RMS error does not oscillate any longer; it is kept fixed in the further iterations. Our experiences also showed that the global minimum for validation sets is usually found very close to the one for the training sets.

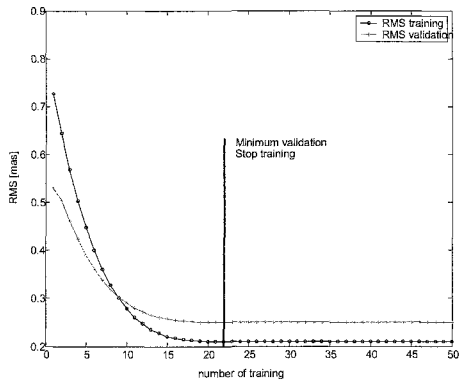


Fig. 3. RMS error during the training. Note that the RMS errors for both training and validation sets are monotonously decreasing.

In numerical examples, for short-term predictions each fuzzy inference system was composed for predicting a particular day between the 1st and the 10th day. After the 10th day, the predicted values were used as inputs for already existing models composed for the previous days' predictions in order to calculate the corresponding prediction values for the days 11, 12,...,40.

Predictions were not extrapolated for further days in future since the RMS error increased rapidly with a linear trend. Instead, individual prediction models for long-term predictions have been formulated to solve this problem. To this end, fuzzy prediction models are composed for ERP in the next 180 days, 270 days and 360 days, respectively.

The error measure was calculated according to Schuh et al. (2002) as

$$RMS_d = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{l}_d^i - l_d^i\right)^2} \qquad (2)$$

where $\hat{l}_d^i$ is the predicted value of the ANFIS network for day d, l_d^i is the actual value of the IERS C04 series, and n is the number of predictions made for the respective day. Approximately 1200 predictions starting at different days in the validation sets of both short and long-term models were made for each day to calculate the RMS error, i.e. n = 1200.

3. 3 Prediction results and comparison

The results of ANFIS prediction were compared with the results of artificial neural networks (ANN) by Schuh et al. (2002) which had performed equally or even better than former methods (Table 1).

Table 1. Comparison of ANFIS prediction with ANN prediction.

	Polar Motion		LOD	
Prediction day	**ANFIS**	**ANN**	**ANFIS**	**ANN**
	RMS^{pm} [mas]	RMS^{pm} [mas]	RMS^{LOD} [ms]	RMS^{LOD} [ms]
1	0.24	0.29	0.017	0.019
2	0.55	0.57	0.045	0.049
3	0.84	0.95	0.067	0.074
4	1.25	1.30	0.088	0.097
5	1.64	1.79	0.115	0.121
10	3.17	3.25	0.188	0.193
15	4.75	4.70	0.251	0.246
20	6.37	6.28	0.259	0.251
25	8.02	7.78	0.267	0.249
30	9.12	8.89	0.275	0.245
35	10.28	10.14	0.281	0.263
40	11.32	10.96	0.290	0.258
180	20.98	23.67	0.296	0.361
270	22.75	24.51	0.313	0.386
360	23.92	25.09	0.303	0.347

Note that the RMS of the ANFIS prediction increases rapidly and it is larger than that of ANN prediction after the 10^{th} day prediction until 40^{th} day in future. This is due to the composition of individual ANFIS network for the prediction of first ten days. Afterwards, the predicted values were used as inputs for the next days to be predicted.

A graphical comparison of the ANFIS results with other methods for the prediction of polar motion regarding the RMS error measure is given in Fig. 4 for short-term predictions. Fig. 5 shows the results of ANFIS predictions of LOD (top) of ten days into the future (black) and the C04 series reduced by the tidal effects and the linear trend function (gray) and the corresponding prediction errors (bottom).

Prediction results for one year in future are given in Fig. 6 for LOD. It can be clearly seen that the difference between the actual LOD series (black) and the predicted LOD series significantly raises after 1998,5. This characteristic was also similar in case of predictions for the 6^{th} and the 9^{th} month in future. Note that reasonable results using the long-term prediction models were only obtained for predictions not more than three years into the future. The user is recommended to update the prediction model with new available data at least for every three years.

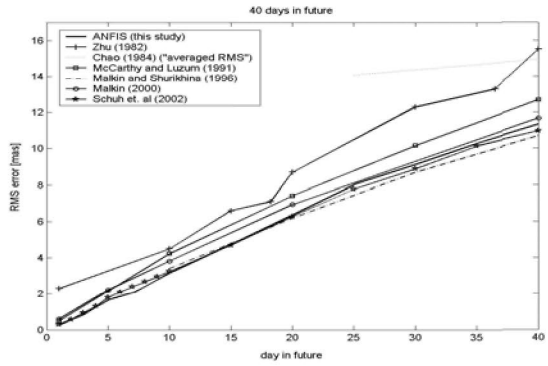


Fig. 4. RMS error of short-term predictions (up to 40 days) of polar motion; ANFIS results and other results according to Schuh et al. (2002)

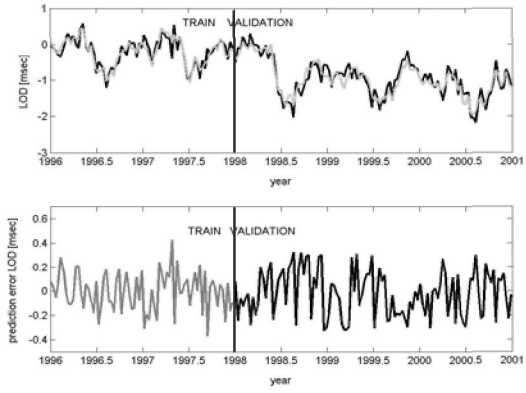


Fig. 5. LOD prediction (*black*) compared with the reduced C04 series (**gray**) and the corresponding prediction errors

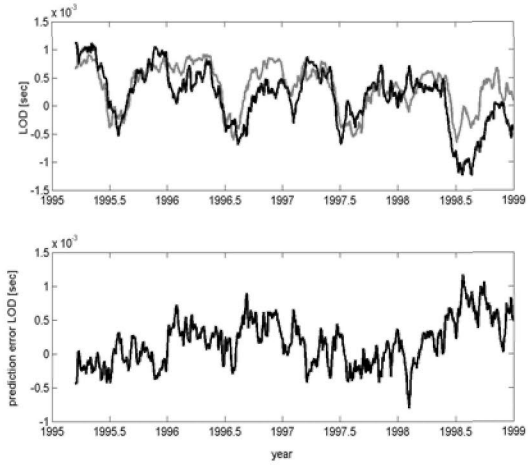


Fig. 6. (Top) Actual LOD series (*black*) and corresponding predictions (*gray*), and prediction errors (bottom)

The one-year prediction results of the y-component of polar motion are given in Fig. 7. Note that the relatively high fluctuations of the prediction errors have a period equal to the one of the Chandler wobble.

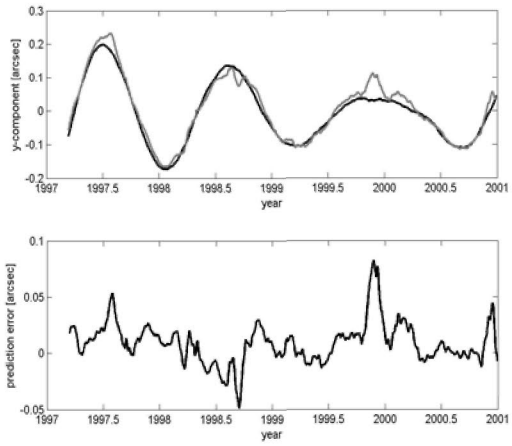


Fig. 7. (top) 1 year prediction results of y-component (*gray*) and corresponding actual time series (*black*), and corresponding prediction errors (bottom)

4 Conclusions

The comparison of ANFIS-derived results with those from other methods clearly indicates that fuzzy inference systems are appropriate tools to predict the ERP. Precise predictions for both the short-term and the long-term period are possible by employing individual adaptive networks for each day of prediction. For predictions of several and more years into the future, more input variables could be used in the fuzzy model because computer run time becomes a minor restriction. However, this case was not considered in this study.

The comparison between the two soft computing methods ANN and ANFIS concerning complexity and the resulting predictions is illustrative. ANN offer a very good prediction but they are rather complex to handle. In addition, one has to calculate a refined a priori model before training the network. This is partially true for ANFIS prediction, but is a less complex task than for the ANN. To our experience, training of ANN takes much longer than of ANFIS: training a feed-forward ANN with two hidden layers and ten inputs may take hours or even days on the same computer used for ANFIS. There are also a smaller number of input variables used in ANFIS prediction models than in ANN prediction models. As a future work, values of the atmospheric and oceanic excitation functions may be added into the ANFIS architecture in order to improve the prediction results.

A promising research can be the use of genetic algorithms for the determination of optimal patterns by evolutionary means. This may allow comparisons between the patterns for different prediction days and may improve the prediction quality.

Acknowledgements

The authors would like to thank DAAD for supporting this project. Special thanks to DGFI (Deutsches Geodätisches Forschungsinstitut) for providing scientific and logistic assistance. The advice of Dr. Katja Heine from Cottbus University of Technology is gratefully acknowledged.

References

Akyilmaz O., Kutterer H. (2003). Prediction of earth rotation parameters by fuzzy inference systems. J. of. Geod. (in review)

Chao B. F. (1984). Predictability of the Earth´s polar motion. Bull. Géod. 59: 81-93

Freedman A. P., Steppe J. A., Dickey J. O., Eubanks T. M., Sung L. Y. (1994). The short-term prediction of universal time and length of day using atmospheric angular momentum. J. of Geophysical Research, Vol. 99, No. B4: 6981-6996

Jang J. S. R. (1993). ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Trans. On Systems, Man and Cybernetics, Vol. 23, pp. 665-685

Malkin Z., Skurikhina E. (1996). On Prediction of EOP. Comm. IAA 93

McCarthy D. D., Luzum B. J. (1991). Prediction of Earth orientation. Bull. Géod. 65: 18-22

McCarthy D. D. (1996). IERS Conventions. IERS Technical Note 21, Paris.

Schuh H., Ulrich M., Egger D., Müller J., Schwegmann W. (2002). Prediction of Earth Orientation Parameters by Artificial Neural Networks. J. of Geodesy, 76: 247-258

Takagi T., Sugeno M. (1983). Derivation of fuzzy control rules from human operator´s control actions. Proc. Of the IFAC Symp. on Fuzzy Information, Knowledge Representation and Decision Analysis, pp. 55-60

Ulrich M. (2000). Vorhersage der Erdrotationsparameter mit Hilfe Neuronaler Netze. IAPG/FESG No. 9, Institut für Astronomische und Physikalische Geodäsie, Technische Universität München

Zadeh L. A. (1965). Fuzzy sets. Information and Control, 8: 338-353

Zadeh L. A. (1996). Foreword in Neuro-Fuzzy and Soft Computing by Jang JSR, Sun CT and Mizutani E, Prentice Hall, ISBN 0-13-261066-3

Zhu S. Y. (1981). Prediction of Earth rotation and polar motion. Reports of the Department of Geodetic Science and Surveying, Report No. 320, Columbus, Ohio

Zhu S. Y. (1982). Prediction of polar motion. Bull Géod. 56: 258-273

SYMPOSIUM G06

Insight into Earth System Science: Variations in the Earth's Rotation and its Gravity Field

C.R. Wilson

Oceanic excitation of polar motion from intraseasonal to decadal periods

Aleksander Brzeziński, Jolanta Nastula, Barbara Kołaczek
Space Research Centre, Polish Academy of Sciences, Bartycka 18A, 00-716 Warsaw, Poland

Rui M. Ponte
Atmospheric and Environmental Research, Inc., 131 Hartwell Ave., Lexington, MA 02421-3126 USA

Abstract. We study the oceanic excitation of polar motion by using six different time series of the non-tidal oceanic angular momentum which are either publicly available or more experimental in nature. The oceanic excitation series are compared to each other and to the difference between the excitation inferred from the polar motion data and the corresponding atmospheric contributions. Comparisons show good agreement between the observed residual excitation and most of the ocean series. By performing computations separately for the seasonal sinusoids and four non-overlapping spectral components (decadal, interannual, seasonal and intraseasonal) we demonstrate that the quality of the ocean products depends on the considered frequency band. We also discuss the importance of using data-constrained ocean models in future efforts to improve ocean angular momentum estimates.

Keywords. Polar motion, oceanic excitation, atmospheric excitation, oceanic angular momentum.

1 Introduction

Recent estimates of non-tidal oceanic angular momentum (OAM) demonstrate the importance of ocean dynamics in driving polar motion at intraseasonal, seasonal and interannual periods; see, e.g., (Nastula and Ponte, 1999) and the recent review by Brzeziński (2003). Adding OAM to atmospheric angular momentum (AAM) signals brings the modeled excitation closer to the observed; nevertheless the remaining discrepancy is still significant. In a previous work (Brzeziński et al., 2002) we discovered that the estimation of the oceanic excitation of the Chandler wobble depends on the OAM series used. One can expect that the differences between various OAM series can be also important for the excitation balance of the other components of polar motion, including the seasonal wobbles, and intraseasonal, interannual and decadal variations.

Here we investigate this problem using various available OAM estimates. The OAM series are compared to each other and to the difference between the excitation inferred from the polar motion data, usually referred to as "observed" or "geodetic" excitation, and the corresponding atmospheric excitation series. An important purpose of this work is to study the dependence of OAM results on the ocean general circulation model (OGCM) and data assimilation procedures used. Comparison with the polar motion data should serve as a useful tool in verifying the overall quality of the ocean products and checking the reliability of the OAM data.

2 Data sets and procedures

We used six different OAM time series. Three of them are available from the website of the International Earth Rotation Service (IERS) Special Bureau for the Oceans (SBO) (IERS, 2002):

P98 Ponte et al. (1998); MIT; NCEP; 5 days; 1985.0 to 1996.3.

J01 Johnson et al. (1999); POCM version 4B; ECMWF; 3 days; 1988.0 to 1998.0.

G03 Gross et al. (2003); ECCO-JPL which is modification of the MIT model; NCEP; 1 day; 1980.0 to 2002.2.

Here and below we use the following description of the OAM series: major reference, model details, forcing, sampling interval, data span. More detailed description of these series and the underlying OGCM's is provided together with the data sets.

The other OAM estimates are the following experimental series:

P01 Ponte (2001); MOM plus data assimilation, developed by J. Carton and colleagues at the University of Maryland; COADS (NCEP for last few years); 1 month; 1950.0 to 2000.0.

Pb Ponte and Ali (2002); barotropic; NCEP; 1 hour; 1993.0 to 2000.5.

Pa Ponte et al. (2001); ECCO-Scripps model (MIT plus data assimilation); NCEP; 1 day; 1992.8 to 1998.0.

Except for Pb, all products are based on full OGCMs that include baroclinic processes. Unmodeled volume changes are corrected for as described by Greatbatch (1994).

We also used the following AAM and polar motion (PM) series:

AAM Salstein and Rosen (1997); products of the NCEP/NCAR reanalysis project (Kalnay et al., 1996); 6 hours; 1948 to 2003.

PM combined solution COMB2002 of Gross (2003), with 12-hourly sampling and spanning 1962 to 2003.

As a first step of analysis, we computed the observed excitation by applying to the PM series the Kalman filter developed by Brzeziński (1992). Then we computed for all the OAM, AAM and observed excitation series the unweighed least-squares (LSQ) fit of the model comprising the 1-st order polynomial and a sum of complex sinusoids with periods ± 1, $\pm 1/2$ and $\pm 1/3$ years. Parameters of the annual, semiannual and terannual terms are used in Sec. 3.1 for studying the excitation balance of polar motion.

For the rest of analysis presented here, we removed the seasonal sinusoids and linear trend from each time series. Next, by applying the standard band-pass filtering we decomposed the excitation series into four non-overlapping components: decadal – periods longer than 10 years, interannual – periods between 2 and 10 years, seasonal – periods between 1/2 and 2 years and intraseasonal – periods between 20 days and 1/2 year. In each band, we compared the OAM series to each other and to the difference between the observed and atmospheric excitation (sections 3.2 to 3.4).

3 Results

3.1 Annual, semiannual and terannual wobbles

Oceanic and atmospheric excitations of the seasonal wobbles are compared to the observed excitations in Figures 1 to 3. Each seasonal elliptical wave is shown as combination of two circular motions with opposite directions, retrograde (clockwise, negative frequency) and prograde (counterclockwise, positive frequency), rather than the usual projection on the x and y axes of the Earth-fixed system. This way of presentation facilitates the physical interpretation of results because the transfer function between the excitation function and polar motion is asymmetrical with respect to zero frequency. For periods $+1$, -1, $+1/2$, $-1/2$, $+1/3$, $-1/3$ years the magnitude of the transfer function is 5.37, 0.46, 0.73, 0.30, 0.39, 0.22, respectively.

This variability of the transfer function also implies that the observed excitations depicted in phasor diagrams are not equally well constrained by geodetic observations of polar motion, as illustrated by the error circles in Figs. 1-3. Their radii were computed by taking the formal standard errors of the ampli-

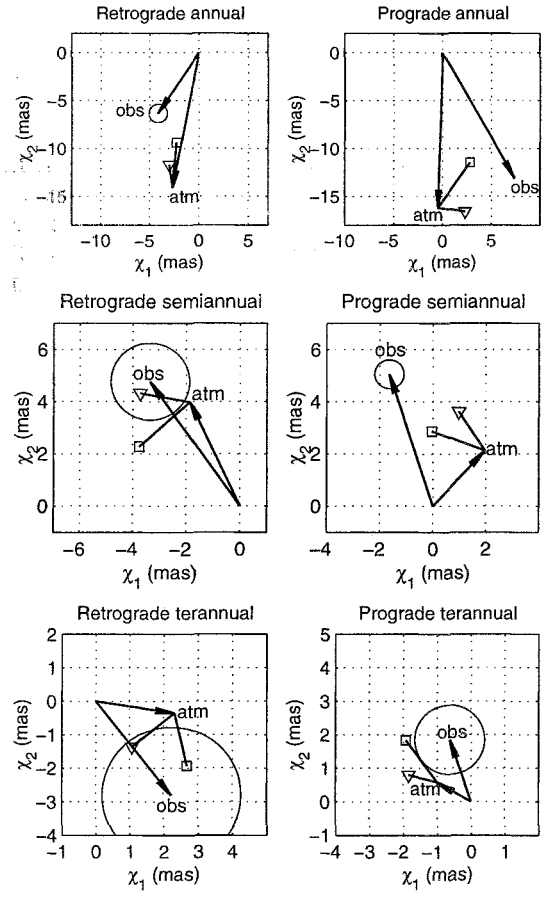


Fig. 1 Phasor diagrams of the seasonal components of the polar motion excitation functions, observed (obs), atmospheric (atm) and oceanic: G03 (square) and P01 (triangle). The atmospheric estimates include the effects of both winds and surface pressure with the IB correction, the oceanic estimates include the effects of both currents and ocean-bottom pressure. Analysis is done over the period 1980.0 to 2000.0. Phase refers to the epoch J2000.0.

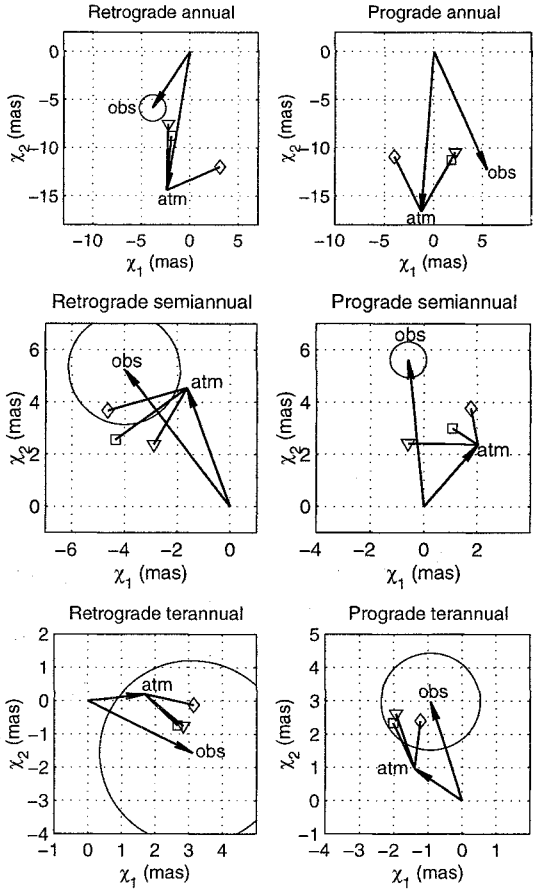


Fig. 2 Same as in Figure 1, but for the following OAM series: P98 (triangle), G03 (square) and J01 (diamond) and for the period 1985.0 to 1996.3.

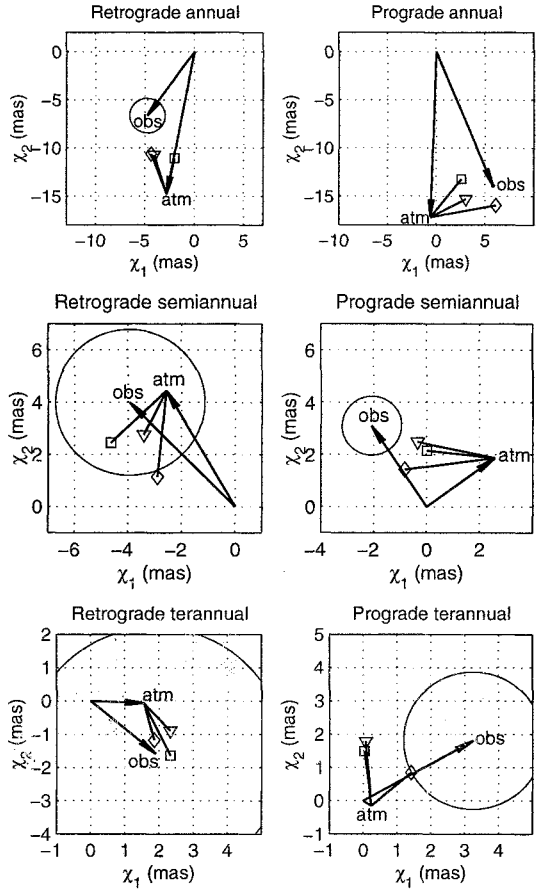


Fig. 3 Same as in Figure 1, but for the OAM series: Pb (triangle), Pa (diamond) and G03 (square) and for the periods 1993.0 to 2000.5 (Pb and G03) and 1992.8 to 1998.0 (Pa).

tudes in the LSQ adjustment, which depend on the length of the input series but have similar values for all six harmonics included in the model, and dividing them by the magnitude of the transfer function. After such scaling the ratio of the observed amplitude to the computed error shown in diagrams, roughly expresses the signal-to-noise ratio in the corresponding waves in polar motion.

In this comparative analysis we divided the OAM series into three groups based mostly of the length of the OAM data. Note, however, that in all three cases we included the series G03 with the purpose of checking how the estimated parameters depend on the period of analysis.

From Fig. 1 it can be seen that the oceanic excitation of seasonal polar motions estimated from the 2-decades long OAM series P01 and G03 shows differences in both magnitude (up to a factor of 2) and phase (up to 60°). Nevertheless, in most cases adding the OAM to the AAM improves considerably the excitation balance.

Comparison of the parameters estimated from the three series P98, J01 and G03 available from the IERS SBO (Fig. 2) shows very good agreement between P98 and G03. Parameters corresponding to J01 roughly agree in magnitude with the other two sets, but there is rather large difference in phase. This difference makes much worse the excitation balance for the annual waves which are the most important components of the comparison. Before interpreting the detected discrepancy one should take into account that P98 and G03 were estimated from the similar ocean models and using the same atmospheric forcing which is consistent with the AAM data adopted in this work.

Comparison of the experimental models Pa and

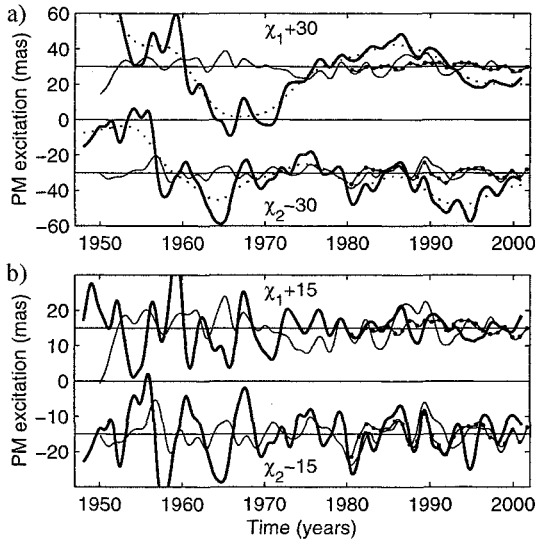


Fig. 4 Oceanic excitation of polar motion at interannual and decadal periods. Comparison is done for the OAM series P01 (thin line), G03 (thin line with dots) and the differences between observed and atmospheric excitations: a) original series (thick line) and its decadal component (dotted line); b) interannual component (thick line).

Pb with the series G03 (Fig. 3) shows considerably good agreement between three sets of the estimated excitation parameters. With exception of the retrograde semiannual term, the best agreement with the observed excitation is obtained for Pa which is the only series in this comparison with data assimilation. This confirms the importance of data assimilation for modeling the ocean circulation and estimating OAM.

3.2 Interannual and decadal variation

There are only two OAM series which could be used for studying the long periodic excitation of polar motion, the 50-year series P01 and the 22-year series G03. These series are compared to the difference between the observed and atmospheric excitations (hereafter denoted as NAM) in Fig. 4a. The comparison confirms earlier conclusions of Brzeziński et al. (2002) and Gross et al. (2003) that the large decadal variation, presumably corresponding to a 30-year "Markowitz wobble", in the NAM series (dotted line) cannot be explained by the oceanic excitation.

Hence, in Fig. 4b we compared the OAM curves to the interannual component of the NAM series. In addition, we computed the correlation between each of the OAM series and the NAM series (Tab. 1), separately for the χ_1 and χ_2 components and for the complex combination $\chi_1 + i\chi_2$. We also computed the percentage of variance reduction in the NAM series when subtracting the oceanic excitation. This quantity is usually positive but takes on negative values if the variance increases. (Similar analysis will be done for the other spectral bands; see secs. 3.3 and 3.4.)

Table 1 Correlation analysis between time series expressing the oceanic excitation of polar motion and the difference between observed and atmospheric excitations, shown in Figure 4b. Complex correlation is expressed by its magnitude and argument. The last two columns show the reduction of variance when subtracting the OAM series from the reference NAM series.

Series	Correlation				Var. red. (%)	
	χ_1	χ_2	$\chi_1 + i\chi_2$		χ_1	χ_2
Figure 4b, 1980.0 $\longrightarrow$ 2000.0						
P01	0.216	0.439	0.366	21.4^o	−158.8	9.7
G03	−0.080	0.518	0.399	-12.0^o	−43.7	25.1

Figure 4b shows that the interannual variation in polar motion decreases considerably with the advent of space geodesy observations. That indicates that the interannual variation observed in polar motion at earlier times may express in part systematic errors in the optical astrometry. During 1980–2000, interannual variation is similar in P01 and G03, and both these OAM series are quite well correlated with the NAM data (Tab. 1), but only for the χ_2 -component.

3.3 Seasonal band

As defined in Sec. 2, by seasonal component of the polar motion excitation we mean here variations with periods between $1/2$ and 2 years, excluding the

Table 2 Same as in Table 1 but for periods between 1/2 and 2 years and time series depicted in Figure 5.

Series	Correlation				Var. red. (%)	
	χ_1	χ_2	$\chi_1 + i\chi_2$		χ_1	χ_2
Figure 5a, 1985.0 $\longrightarrow$ 1996.3						
P98	0.477	0.665	0.605	-7.9^o	9.7	38.5
G03	0.561	0.738	0.682	-8.1^o	25.5	53.1
Figure 5b, 1988.0 $\longrightarrow$ 1997.9						
P01	0.629	0.351	0.466	-0.6^o	8.2	−3.0
J01	0.455	0.316	0.357	-3.1^o	12.3	−27.9
Figure 5c, 1993.0 $\longrightarrow$ 1998.0						
Pb	0.523	0.579	0.562	-5.4^o	27.1	32.9
Pa	−0.004	0.546	0.346	-0.2^o	−35.3	29.8

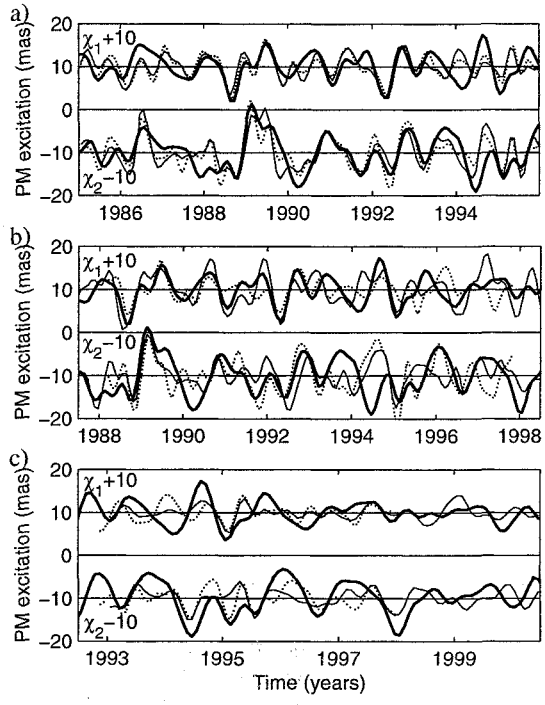


Fig. 5 Oceanic excitation of polar motion at periods between 1/2 and 2 years. Compared are the OAM series:
a) P98 (thin line) and G03 (dotted line),
b) P01 (thin line) and J01 (dotted line),
c) Pb (thin line) and Pa (dotted line).
The oceanic contribution is compared to the difference between observed and atmospheric excitations (thick line).

sinusoidal terms considered separately in Sec. 3.1. Hence, this component contains all departures of the seasonal excitation from the purely sinusoidal pattern and, what is particularly important, the whole contribution to the excitation of the free 14-month Chandler wobble. The oceanic contribution to the excitation of the Chandler wobble could not be consider here separately due to the limited space. Instead, we will perform in the seasonal band only the overall comparative analysis, by applying the same method as in the other frequency bands.

Fig. 5a shows that, as in case of the parameters of the seasonal sinusoid, there is good agreement in the seasonal band between P98 and G03 data which can be attributed to the similarity of the underlying OGCMs. From Tab. 2 these two series are also quite well correlated with the NAM series. Better results for G03 are probably related to the higher model resolution in both the horizontal and the vertical.

In case of the series P01 and J01 depicted in Fig. 5b, the overall agreement with the NAM series is lower than for P98 and G03, particularly for the χ_2 -component; see Tab. 2. In case of the experimental OAM series Pb and Pa (Fig. 5c, Tab. 2) the agreement with the observed excitation is rather good with one exception, namely the anomalous behaviour of the χ_1 -component of Pa.

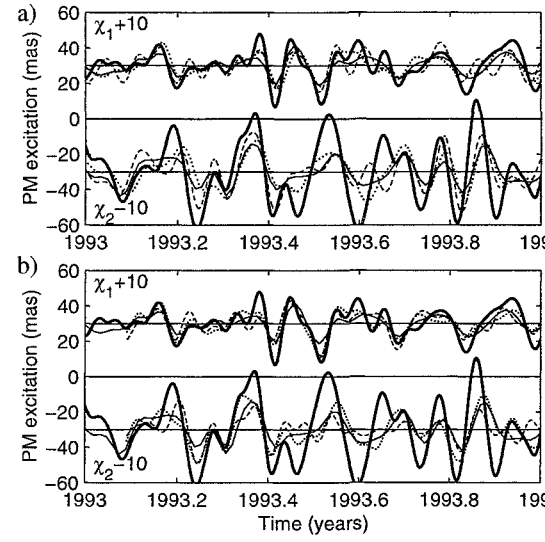


Fig. 6 Oceanic excitation of polar motion at intraseasonal periods. Compared are the OAM series: G03 (thin line),
a) P98 (dashed line), J01 (dotted line);
b) Pb (dashed line), Pa (dotted line).
The oceanic contribution is compared to the differences between observed and atmospheric excitations (thick line).

3.4 Intraseasonal band

Comparison done in Fig. 6a shows that in the intraseasonal band all the IERS SBO series (P98, G03 and J01) are similar to each other and, from Tab. 3, are also quite well correlated with the reference

Table 3 Same as in Table 1 but for intraseasonal periods and time series depicted in Figure 6.

Series	Correlation				Var. red. (%)	
	χ_1	χ_2	$\chi_1 + i\chi_2$		χ_1	χ_2
Figure 6a, 1989.2 ⟶ 1995.9						
P98	0.605	0.575	0.587	$-7.8°$	36.2	33.0
G03	0.658	0.619	0.638	$-10.3°$	41.4	34.7
J01	0.596	0.608	0.617	$-11.0°$	35.5	36.7
Figure 6b, 1993.1 ⟶ 1997.7						
Pb	0.700	0.692	0.678	$2.4°$	49.0	41.3
Pa	0.773	0.722	0.731	$-3.8°$	59.7	48.7

NAM series. Again, the best results are obtained for G03 which confirm a high reliability of this OAM series. Even better results are obtained for the experimental series Pb and Pa (Fig. 6b, Tab. 3). The highest correlation with the observed excitation and the highest reduction of variance in the observed excitation, is for Pa, the OAM series estimated from the output fields of the OGCM with data assimilation.

4 Summary and conclusions

Our results prove that adding the oceanic excitation of polar motion to the atmospheric excitation brings the modeled excitation closer to the observed in most of the spectral bands considered here. The only exception is the decadal component, where the large variation in the observed excitation, presumably corresponding to a 30-year "Markowitz wobble", cannot be explained by the combination of the atmospheric and oceanic processes when considering the available data and models.

Comparisons show a close relationship between the excitation estimates and the underlying ocean circulation models. An important observation is that the quality of the ocean analyses depends on the considered frequency band. In the seasonal band, for example, the POCM-based series J01 gives considerably worse agreement with the PM and AAM data than the MIT-based series P98 and G03. Nevertheless, at intraseasonal periods all these three OAM series yield comparable results. As another example, one can compare the correlation coefficients between Pa and NAM in the seasonal and intraseasonal bands.

Among the three series which are available from the IERS SBO, that is P98, J01 and G03, the best agreement with the observed minus atmospheric excitation has been obtained for G03. Regarding the benefits of data assimilation, Pa appeared to be the best series at intraseasonal periods, but no particular gain is seen in P01. The impact of data assimilation should continue to be addressed in future work.

Acknowledgments

This work has been supported by the Polish Ministry of Scientific Research and Information Technology under grant No. 5 T12E 039 24. J.N. expresses her gratitude for the travel grant provided by the IUGG2003 LOC and cosponsoring Associations. We thank C. Ron for providing the LaTeXpackage to format the paper according to IAG guidelines and M. Kalarus for his help in converting the figures to the black/white version.

References

Brzeziński, A. (1992). Polar motion excitation by variations of the effective angular momentum function: considerations concerning deconvolution problem. *manuscripta geodaetica*, vol. 17, pp. 3–20.

Brzeziński, A. (2003). Oceanic excitation of polar motion and nutation: an overview. In: *IERS Technical Note No. 30*, (ed.) B. Richter. Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, in press.

Brzeziński, A., J. Nastula and R. M. Ponte (2002). Oceanic excitation of the Chandler wobble using a 50-year time series of ocean angular momentum. In: *Vistas for Geodesy in the New Millennium*, (eds.) Ádám and Schwarz. IAG Symposia, Vol. 125, Springer Verlag, New York.

Greatbatch, R. J. (1994). A note on the representation of steric sea level in models that conserve volume rather than mass. *J. Geophys. Res.*, vol. 99, pp. 12,767–12,771.

Gross, R. S. (2003). Combinations of Earth Orientation Measurements: SPACE2002, COMB2002 and POLE2002. Tech. Rep. JPL Publication 03-011, JPL, Pasadena, CA.

Gross, R. S., I. Fukumori and D. Menemenlis (2003). Atmospheric and oceanic excitation of the Earth's wobble during 1980-2000. *J. Geophys. Res.*, p. in press.

IERS (2002). IERS Annual Report 2001. Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main.

Johnson, T. J., C. R. Wilson and B. F. Chao (1999). Oceanic angular momentum variability estimated from the Parallel Ocean Climate Model, 1988-1998. *J. Geophys. Res.*, vol. 104, pp. 25,183–25,195.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins and et al. (1996). The NMC/NCAR 40-year reanalysis project. *Bull. Amer. Met. Soc.*, vol. 77, pp. 437–471.

Nastula, J. and R. M. Ponte (1999). Further evidence for oceanic excitation of polar motion. *Geophys. J. Int.*, vol. 139, pp. 123–130.

Ponte, R. M. (2001). A 50-year time series of ocean angular momentum for Earth rotation studies. Unpublished manuscript.

Ponte, R. M. and A. H. Ali (2002). Rapid ocean signals in polar motion and length of day. *Geophys. Res. Lett.*, vol. 29, p. 10.10129/2002GL015312.

Ponte, R. M., D. Stammer and J. Marshall (1998). Oceanic signals in observed motions of the Earth's pole of rotation. *Nature*, vol. 391, pp. 476–479.

Ponte, R. M., D. Stammer and C. Wunsch (2001). Improving ocean angular momentum estimates using a model constrained by data. *Geophys. Res. Lett.*, vol. 28, pp. 1775–1778.

Salstein, D. A. and D. Rosen (1997). Global momentum and energy signals from reanalysis systems. In: *Proc. 7th Conference on Climate Variations, 2-7 February 1997, Long Beach, California*. American Met. Soc., Boston, Massachusetts.

EXCITATIONS OF POLAR MOTION FROM AN ENSEMBLE OF GLOBAL ATMOSPHERIC MODELS

J. Nastula [1], D. Salstein [2] and B. Kolaczek [1]

(1) Space Research Center of the Polish Academy of Sciences;
(2) Atmospheric and Environmental Research, Inc.

Abstract We study how atmospheric models can be best used to understand excitations of polar motion. The set of models consists of those contributed worldwide by meteorological centers participating in the Atmospheric Model Intercomparison Project, which are forced solely by a prescribed distribution of sea surface temperatures. We have calculated the monthly mean values of excitations from the resulting surface pressure fields over the 17-year study period. With such historical variability and with the spread inherent in the suite of models, a measure of the excitations of polar motion can be obtained, including variability at intraseasonal and seasonal scales; from these models we can also view the resulting interannual variability, including the impact of El Nino events. We compare the results with the excitations determined by atmospheric analyses, as well as with the geodetic excitation functions themselves, determined from analysis of observations. Based on the suite of atmospheric models we estimate the spread in the polar motion excitation from the ensemble of 19 different models.

Keywords. Atmospheric models, polar motion excitation.

1 Introduction

In this study we use atmospheric models to best understand excitations of polar motion. The set of models is derived from the second phase of the Atmospheric Model Intercomparison Project (AMIP-2; Gates et al. 1999). These models, which cover the 17-year period 1979-1995, are forced solely by a consistent set of monthly mean sea surface temperatures. At present, 20 models have output for AMIP-2, 19 of which can be used for this polar motion study. The models are listed in Table 1. The models have variable resolutions, techniques for numerical computations, as well as a number of physical differences including cloud descriptions, radiative transfer algorithms, and land-surface schemes that involve heat and moisture transfer from the ground.

Results from AMIP are used to determine the best attributes of modeling so that models can be used to project the possible atmospheric state into the future for the examination of global change issues. To evaluate the models in a careful way, a number of Diagnostic Subprojects were organized by the community. The present study falls in the context of the Subproject examining atmospheric angular momentum and its relationship to Earth rotation. Following the study of the axial component of angular momentum and its relationship with length of day in the AMIP models (Hide et al. 1997, Salstein et al. 2003), we expand the subject by focusing here on the equatorial component, related to excitations of polar motion.

We intercompare all the model results to determine their spread, and we also compare the results with analyses of polar motion excitation using the surface pressure fields from the NCEP-NCAR Reanalysis (Kalnay et al. 1996). To complement the analysis we also use the geodetic measure of polar motion excitation, derived from the combined set of Earth orientation parameters (EOP C03; IERS, 2002). This comparison is novel for the AMIP project because it involves a data set that is entirely independent of the atmosphere. As in other AMIP comparisons, we examine variability at both subseasonal and seasonal time scales.

2 Computations

Using the monthly mean values of surface pressure from the AMIP models, we compute the excitation functions χ_1 and χ_2 about the axes 0-180 and 90E-90W, respectively, according to the Barnes et al. (1983) formulation. We consider our results for two cases, first when polar motion excitation is dependent primarily on atmospheric surface pressure (P), and second when the atmospheric over the oceans acts as an Inverted Barometer (P + IB). The additional wind term, though present in the excitations, is of lesser importance, and it is not treated here in this first analysis. Similar excitations were derived from the NCEP-NCAR reanalysis data. We also computed the excitation function from the geodetic polar motion

data according to an inverse solution (Barnes et al., 1983). Because of the lower agreement in the earlier years between the NCEP-NCAR reanalysis and the geodetic data (Nastula and Salstein 1999), we limit that aspect of the analysis to the period since the early 1980s.

3 Results

The 19 AMIP models were first averaged according to calendar months so we can estimate the seasonal cycle ensemble, including its spread. The two components χ_1 and χ_2 have cycles both with a maximum around June/July (Fig. 1). The cycle is larger in χ_2 than χ_1 for the pressure term, for which the seasonal differences are 3 and 5 x 10^{-7} rad respectively. Introduction of the IB reduces amplitudes of the seasonal cycle in both cases, to 1 and 3 x 10^{-7} rad respectively. While the NCEP-NCAR results was found to be within the spread of models in nearly all cases, a measure of the quality of the model results may be obtained by computing residuals of each model from the NCEP-NCAR series (Fig. 2a and 2b, for P and P+IB respectively). For the P term, the residuals in χ_1 are smaller than χ_2. The residual are reduced further by introduction of the P+IB, especially for the χ_1 term, reduced by a factor of between 2 and 3. For the P terms the higher residuals are in models 1, 4, 13, 16 and 17; in the P+IB case, they are in models 1, 10, 13, 16, and 18. The smallest residuals are in models 3, 5, 9, 12, and 15 from which 12 and 15 contained the Land Surface Scheme of Bonan et al. (1996).

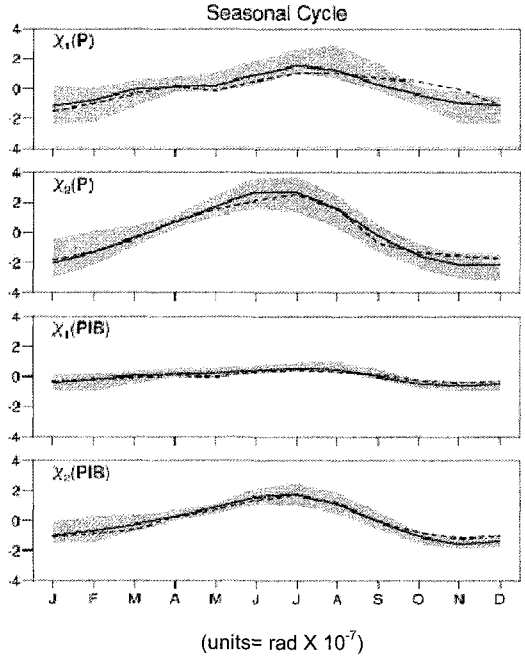


Fig. 1. Ensemble mean value (solid line) and spread (shaded area) at each calendar month of pressure and pressure+IB excitation terms for polar motion and NCEP-NCAR reanalysis values (dashed line.)

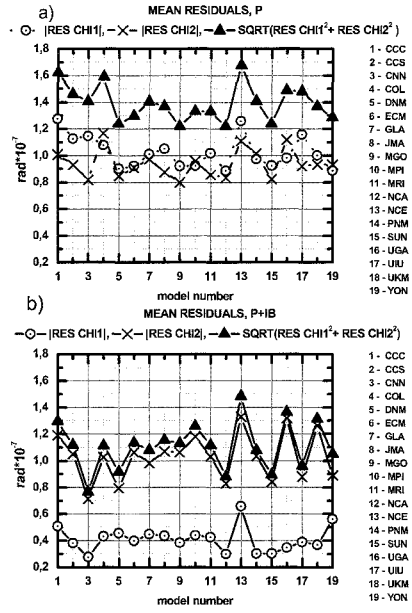


Fig. 2. Residuals of a) pressure terms and b) pressure +IB terms from models used in AMIP-2 from NCEP-NCAR.

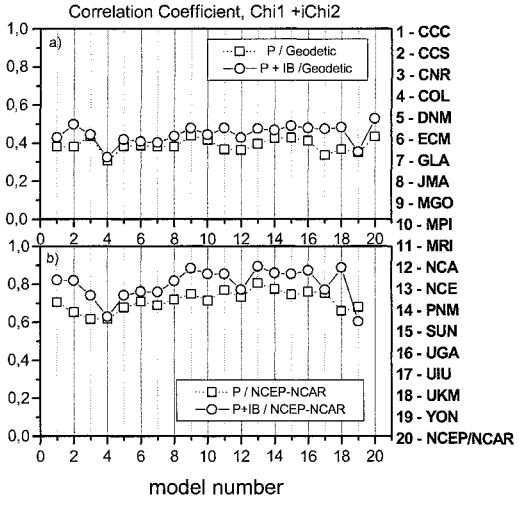


Fig. 3. Correlations of pressure and pressure+IB excitation terms from models used in AMIP with both those of NCEP-NCAR and geodetic excitations of polar motion.

4 Correlation of model results with NCEP-NCAR reanalysis and geodetic data

The model polar motion excitations are correlated with those of the geodetic excitations and the NCEP-NCAR reanalysis observations in Fig. 3. The complex combination of components is shown here. The model results are correlated to reanalysis data on the level between 0.6 and 0.9, surpassing the correlations with geodetic data, which fall between levels of 0.4 to 0.6. The difference is not surprising considering that factors other than the atmosphere, like the ocean and hydrology, are known to play a role in polar motion. In addition, the P+IB analyses have better correlations with geodetic results than do the P results (with one exception, model 19). Also, the P+IB results from models are better correlated with P+IB from NCEP-NCAR than are the corresponding P results, though there are several exceptions to this pattern too. Models 4 and 19 have generally the lowest correlations here. In the top panel of the figure, line 20 shows the correlation between observational NCEP-NCAR and geodetic data; better results occur with P+IB than P for this comparison too. Interestingly, for the P case, model data are sometimes better correlated with the geodetic excitations than is NCEP-NCAR.

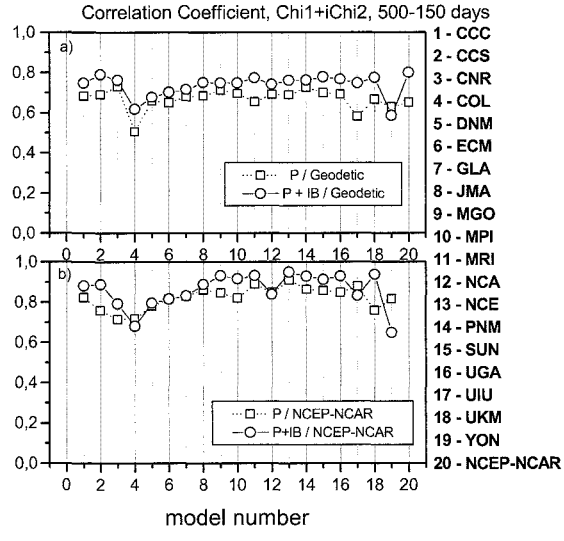


Fig. 4. Correlations of pressure and pressure+IB excitation terms from models used in AMIP over the temporal spectral range 150-500 days with both those of NCEP-NCAR and geodetic excitations of polar motion.

It is known that correlation between the atmosphere and the geodetic excitations of polar motion may be related to the time scale at which they occur, so we extract the seasonal band between 500 and 150 days (Fig. 4). We note that the correlations between models and geodetic signals in this band are now in the 0.6 to 0.8 range. The time variability of this relationship for that band is shown in Fig. 5. Here it is apparent that there is a time relationship in which most of the models have a lower correlation with both NCEP-NCAR and the geodetic results during certain epochs, for example 1983-84 and 1990. For the first epoch, we speculate that the low value may be related to the strong El Nino/La Nina sequence of events then, for which the ocean may be playing a stronger role than usual. The subseasonal band has considerably lower correlations overall, even lower than 0.2, so the comparable figures are not shown here. The monthly resolution of the model data may be limiting the possible correlations.

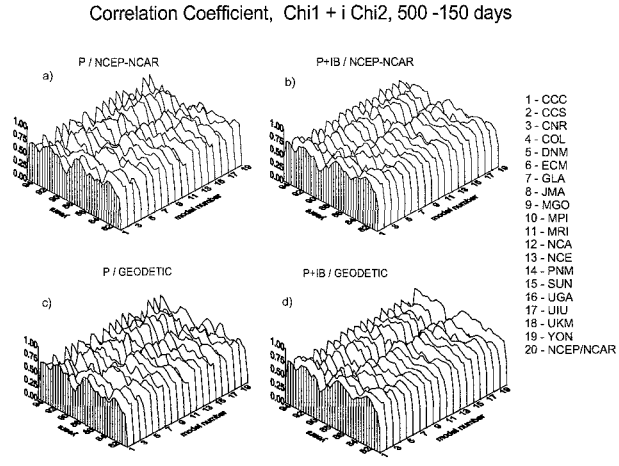


Fig. 5. Correlations of pressure and pressure-excitation terms from 19 models used in AMIP by time period. (Upper left—P vs. NCEP/NCAR; Upper right P+IB vs. NCEP/NCAR; lower left-P vs. geodetic; lower right—P+IB vs geodetic.) Years are 1981-1995, from the front.

5 Conclusions

We have intercompared AMIP-2 model polar motion excitation results and compared them against those of a meteorological analysis system and geodetically-determined results. We find that on seasonal scales, each AMIP model is good enough to be used for polar motion analyses, but on subseasonal scales, they are typically not well correlated with polar motion. There are certain epochs for which residuals of model results and both observations and geodetic results are large, related in part to El Nino events.

Models do behave differently in both time scales and have a spread of results. A purpose of AMIP is to

determine which are the model attributes of special importance for simulations of many aspects of atmospheric behavior, and the best way of defining these attributes. Here we have noted that the land-surface scheme appears to play some role in the behavior, and that some models with the smallest residuals, for example, contain the Land Surface Scheme of Bonan et al. (1996) (Table 1). Other models may do well too, including successful cloud physics and other physical parameters.

Acknowledgments The work of JN and BK is sponsored by the Polish Committee for Scientific Research, and that of DS is sponsored by the US National Science Foundation and NASA Solid Earth and Natural Hazards Program. We thank the Atmospheric Model Intercomparison Project at DOE's Lawrence Livermore Laboratory for making the data available. JN received a travel grant from the IUGG and co-sponsoring organizations.

References

Barnes, R.T.H., R. Hide, A.A. White, and C.A. Wilson 1983: Atmospheric angular momentum fluctuations, length-of-day changes and polar motion, *Proc. R.Soc. Lond.*, **A387**, 31-73.

Bonan, G., 1998: The land surface climatology of the NCAR land surface model coupled to the NCAR Community Climate model. *J. Climate*, **11**, 1307-1326..

Gates, W. L., J. Boyle, C. Covey, C. Dease, C. Doutriaux, R. Drach, M. Fiorino, P. Gleckler, J. Hnilo, S. Marlais, T. Phillips, G. Potter, B. Santer, K. Sperber, K. Taylor, and D. Williams, 1998: An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I), *Bull. Amer. Meteorol. Soc.*, **73**, 1962-1970.

Hide, R., J.O. Dickey, S.L Marcus, R.D. Rosen, and D.A. Salstein, 1997: Atmospheric angular momentum fluctuations in global circulation models during the period 1979-1988. *J. Geophys. Res.*, **102**, 16423-16438.

IERS, 2002: The International Earth Rotation Service Annual Report.

Kalnay et al., 1996: The NCEP/NCAR 40-year Reanalysis Project, *Bull. Amer. Meteor. Soc.*, **77**, 437-41.

Nastula, J. and D.A. Salstein, 1999: Regional atmospheric momentum contributions to polar motion, *J. Geophys. Res.*, **104**, 7347-7358

Salstein, D.A., R.D. Rosen, J.O. Dickey, and S.L Marcus, 2003: Assessment of atmospheric angular momentum in a range of general circulation models, manuscript in preparation.

Table 1. AMIP models

Model	Modeling center	Resolution	Land surface scheme
SUNYA	State University of New York at Albany	T42 L18	LSM Bonan scheme
COLA	Center for Ocean-Land-Atmosphere Studies	R40 L18	SiB
UKMO	U.K. Meteorological Office	3.75x2.5 L19	MOSES
CCC	Canadian Centre for Climate Research	T47 L32	"bucket"
CCSR	Center for Climate System Research	T42 L18	"bucket"
CNRM	Centre Nationale de Recherches Meteorologiques, France	T63 L45	ISBA
DNM	Department of Numerical Mathematics of the Russian Academy of Sciences	4x5 L21	Volodin and Lykossov scheme
ECMWF	European Centre for Medium-Range Weather Forecasts	T63 L50	Blondin and Böttgerm scheme
GISS	Goddard Institute for Space Studies	4x5 L12	Abramopoulos et al. scheme
GLA	Goddard Laboratory for Atmospheres	4x5 L20	SiB
JMA	Japan Meteorological Agency	T63 L30	SiB
MGO	Main Geophysical Observatory	T30 L14	other
MPI	Max Planck-Institut für Meteorologie	T42 L19	"bucket"
MRI	Meteorological Research Institute, Japan	T42 L30	SiB
NCAR	National Center for Atmospheric Research	T42 L18	LSM Bonan scheme
NCEP	National Center for Environmental Prediction	T42 L18	"bucket"
PNNL	Pacific Northwest National Laboratory	T42 L18	BATS
UGAMP	Universities Global Atmospheric Modelling Programme, U.K.	3.75x2.5 L58	MOSES
UIUC	University of Illinois at Urbana-Champaign	4x5 L24	other
YONU	Yonsei University, Korea	4x5 L15	"bucket"

Land surface schemes

* The Simple Biosphere (SiB) scheme
* The Meteorological Office Surface Exchange Scheme (MOSES)
* The Land Surface Model (LSM)
* Biosphere-Atmosphere Transfer Scheme (BATS)
* The Interactions between Soil-Biosphere-Atmosphere (ISBA) scheme
* "bucket

Sensitivity Analysis of the Non-Linear Liouville Equation

Florian Seitz, Hansjörg Kutterer
Deutsches Geodätisches Forschungsinstitut (DGFI), Marstallplatz 8, D-80539 Munich, Germany
E-Mail: florian.seitz@dgfi.badw.de

Abstract. The non-linear gyroscopic model DyMEG has been developed at DGFI in order to study the interactions between geophysically and gravitationally induced polar motion and the Earth's free wobbles, in particular the Chandler oscillation. The model is based on a triaxial ellipsoid of inertia. It does not need any explicit information concerning amplitude, phase, and period of the Chandler oscillation. The characteristics of the Earth's free polar motion are reproduced by the model from rheological and geometrical parameters. Therefore, the traditional analytical solution is not applicable, and the Liouville equation is solved numerically as an initial value problem. The gyro is driven by consistent atmospheric and oceanic angular momenta. Mass redistributions influence the free rotation by rotational deformations. In order to assess the dependence of the numerical results on the initial values and rheological or geometrical input parameters like the Love numbers and the Earth's principal moments of inertia, a sensitivity analysis has been performed. The study reveals that the pole tide Love number k_2 is the most critical model parameter. The dependence of the solution on the other mentioned parameters is marginal.

Keywords. Earth Rotation, Gyroscopic Model, Liouville Differential Equation, Critical Parameters, Sensitivity Analysis.

1 Introduction

The non-linear gyroscopic model DyMEG (Dynamic Model for Earth Rotation and Gravity) has been developed at DGFI in order to study the dynamics of the Earth system based on the interactions between its individual components. Among these are the atmosphere, the ocean, and the solid Earth. Besides, the influences of sun and moon are considered.

Mass motions outside and inside the Earth are due to geophysical processes and gravitational influences of celestial bodies. They affect the Earth's rotation on sub-daily to secular time scales by polar motion and length-of-day variations (ΔLOD). These variations are superposed by free wobbles of the Earth such as the Chandler oscillation.

DyMEG is based on the balance of angular momentum in the Earth system. The characteristics of Earth rotation are generated by means of rheological and geometrical parameters. No explicit information about amplitude, phase, and period of the Earth's free polar motion is needed as the Liouville differential equation is solved numerically by means of a Runge-Kutta-Fehlberg method (Seitz and Kutterer, 2002; Seitz et al., 2004). In the present investigation, the sensitivity of the solution is discussed with respect to some model parameters which are entered into the model. In particular, the effects of the pole tide Love number k_2, the initial values for the numerical solution and the modification of the initial tensor of inertia of the basic Earth model are discussed.

2 Configuration of DyMEG

2.1 The Liouville Differential equation

With respect to a terrestrial reference system, the Earth's reaction on mass redistributions can be described by the Liouville differential equation:

$$\frac{\mathrm{d}}{\mathrm{dt}}\left(\underline{\mathbf{I}}\,\boldsymbol{\omega}+\mathbf{h}\right)+\boldsymbol{\omega}\times\left(\underline{\mathbf{I}}\,\boldsymbol{\omega}+\mathbf{h}\right)=\mathbf{L}\,. \qquad (1)$$

The vector $\boldsymbol{\omega}(t)$ denotes the rotation vector of the terrestrial system with respect an inertial system. Temporal variations of the Earth's rotation are interpreted as small deviations from a uniform rotation. In the terrestrial system, the coordinates of the Earth rotation vector are expressed by $\boldsymbol{\omega}(t)=\Omega\cdot\left(m_1(t),m_2(t),1+m_3(t)\right)$ where $\Omega\approx 2\pi/$ 86164 s is the approximate angular velocity of the terrestrial system. Small deviations of the instantaneous Earth's rotation axis from the uniform rotation are denoted by the dimensionless quantities m_i(i = 1,

2, 3). The z-axis of the terrestrial reference system is directed approximately towards the Earth's maximum moment of inertia C, the x-axis points towards the Greenwich meridian and the y-axis towards 90°E.

Geophysical and gravitational forces cause motions of masses in individual system components. They show up as changes of the Earth's tensor of inertia $\underline{\mathbf{I}}(t)$ and angular momentum $\mathbf{h}(t)$ with respect to the rotating reference system. The vector $\mathbf{L}(t)$ on the right-hand side of Eq. (1) denotes torques caused by direct gravitational forces of Sun and Moon.

The tensor of inertia $\underline{\mathbf{I}}(t)$ is composed of two components $\underline{\mathbf{I}}_0$ and $\Delta\underline{\mathbf{I}}(t)$, where $\underline{\mathbf{I}}_0$ is the approximate tensor of inertia of the Earth. With respect to the principal axes of inertia, $\underline{\mathbf{I}}_0$ is given by

$$\underline{\mathbf{I}}_0 = \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix}, \qquad (2)$$

where A and B are the equatorial principal moments of inertia ($C>B>A$). $\Delta\underline{\mathbf{I}}(t)$ contains perturbations of $\underline{\mathbf{I}}_0$ due to mass redistributions. The principal axes of inertia and the axes of the terrestrial reference system do not coincide because the axis of the minimum equatorial moment of inertia A points approximately towards 345° longitude (Marchenko and Schwintzer, 2003). This divergence is taken into account by means of a rotation. The dependence of the numerical solution on the values for A, B, and C and on the orientation of the principal axes of inertia is investigated in this paper.

2.2 Free rotation of the gyroscopic model

In order to study the interactions of forced and free wobbles of the Earth, an analytical solution of the Liouville equation is not possible because the Earth's free polar motion is determined within the gyroscopic model by rheological and geometrical parameters (Seitz et al, 2004).

The free wobble of the Earth is lengthened from the Euler period of 304 days (which would be the period if the Earth were a rigid body) to the observed Chandler period of about 434 days due to the influence of rotational deformations (pole tides). This back-coupling mechanism of rotational variations causes perturbations of the second-degree spherical harmonic geopotential coefficients ΔC_{21} and ΔS_{21} (McCarthy, 2003) which are directly linked to $\Delta\underline{\mathbf{I}}(t)$ (Moritz and Mueller, 1987).

$$\Delta C_{21} = -\frac{\Omega^2 a^3}{3GM}\left(\Re(k_2)\cdot m_1 + \Im(k_2)\cdot m_2\right)$$

$$\Delta S_{21} = -\frac{\Omega^2 a^3}{3GM}\left(\Re(k_2)\cdot m_2 - \Im(k_2)\cdot m_1\right)$$

Here, a denotes the Earth's mean equatorial radius, M is the total mass of the Earth and G is the gravitational constant. The effects on the centrifugal potential due to the rheological characteristics of the model body are described by the complex pole tide Love number k_2, which includes the effects of mantle anelasticity and ocean pole tides. Both period and damping of the Chandler oscillation strongly depend on the value of k_2. Its influence on the numerical results is discussed in section 3.1.

In the following, a simple Earth model is employed. It consists of an anelastic mantle and a spherical liquid core which are assumed to be fully decoupled. For investigations of polar motion, this simplification is justified on time scales which are longer than one day (Brzezinski, 2001). Therefore, the principal moments of inertia A, B, and C in Eq. (2) are replaced by A_m, B_m, and C_m which are attributed to the mantle alone (Sasao et al., 1980).

Initial values $m_i(t = t_0)$ for the first time step are deduced from the geodetically observed time series C04 of the IERS. The transformation from the C04-values which describe the celestial ephemeries pole (CEP) with respect to the IERS reference pole and m_i was given by Gross (1992).

2.3 Atmospheric and Oceanic excitation

DyMEG is driven by variations of the atmospheric and oceanic angular momentum. The indirect effect due to load deformations is computed by Green's functions. For atmospheric and oceanic forcing, two independent consistent model combinations are considered.

First, atmospheric data based on the reanalyses of the National Centers of Environmental Prediction (NCEP) (Kalnay et al., 1996) were applied in combination with the ocean model ECCO (Stammer et al., 2003). The combination NCEP + ECCO is a consistent representation of dynamics and mass motions in the subsystems atmosphere and ocean because NCEP forcing fields are used for the computation of ocean dynamics in ECCO. As atmospheric pressure forcing is excluded, the ocean's response to pressure variations is assumed to be inverse barometric. The simulations cover a range of 23 years from 1980 until 2002.

Second, the atmospheric model ECHAM3-T21 GCM (Roeckner et al., 1992), which is driven by observed sea surface temperature (SST) fields, was

used in combination with the ocean model OMCT for circulation and tides (Thomas et al., 2001) which is driven by ECHAM3. Both models and their coupling are described in detail by Seitz et al. (2004). As atmospheric pressure forcing is taken into account by OMCT, the two considered model combinations differ with respect to the pressure coupling. The ECHAM3 + OMCT simulations cover a range of 22 years from 1973 until 1994.

As the NCEP data set is based on atmospheric observations, the combination NCEP + ECCO is expected to correspond better with reality than ECHAM3 + OMCT. The latter models are completely free. Apart from the initial SST-boundary conditions, the dynamics of the atmosphere and the ocean are solely based on model physics. However, it was shown that the polar motion time series resulting from DyMEG are in good agreement with the observations (Seitz et al., 2004).

3 Sensitivity Analysis

3.1 Effect of the pole tide Love number k_2

In order to assess the influence of the pole tide Love number k_2 on polar motion, several model runs, which differ with respect to the value of k_2, were performed. As DyMEG accounts for the effects of equilibrium ocean pole tides and mantle anelasticity, the effective pole tide Love number k_2 is composed of three components:

$$k_2 = k_2^* + \Delta k_2^O + \Delta k_2^A$$

Here, k_2^* denotes a Love number which would be appropriate for a purely elastic Earth. The dynamic response of the ocean as well as the influence of mantle anelasticity are taken into account by adding the supplementary Love numbers Δk_2^O and Δk_2^A, respectively (Smith and Dahlen, 1980). Dynamic effects of the core are neglected because core and mantle are considered decoupled. In DyMEG, the approximate values $k_2^* = 0.3$, $\Delta k_2^O = 0.044$, and $\Delta k_2^A = 0.012 + 0.0035\mathrm{i}$ (Mathews et al., 2002; McCarthy, 2003) are introduced. As the anelastic response of the mantle on rotational variations is accompanied by energy dissipation, the supplement Δk_2^A is complex. As a consequence, the Chandler wobble is a damped oscillation which would diminish if it was not continuously excited.

In the following, the dependence of the numerical solution on the real and imaginary parts of k_2 is studied separately. First, $\Re(k_2)$ was varied from 0.310 to 0.380 in 35 equidistant steps of 0.002 while $\Im(k_2) = 0.0035$ was kept unchanged. DyMEG was driven by atmospheric and oceanic excitation (NCEP + ECCO) including loading and tidal deformations between 1980 and 2002. In Fig. 1 the resulting time series of the x-component of polar motion are shown for the different values of $\Re(k_2)$. As both annual and Chandler wobbles are almost circular, the y-component looks similar.

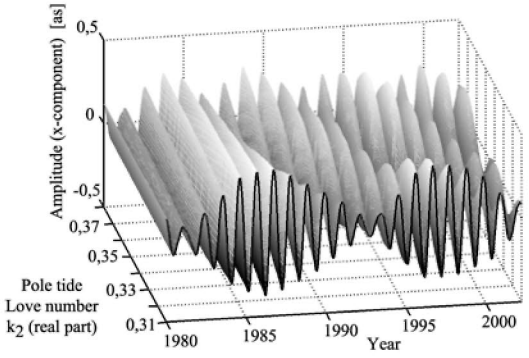


Fig. 1: Resulting time series of polar motion (x-component) for NCEP + ECCO forcing, using different values of $\Re(k_2)$.

As clearly visible, the real part of k_2 has an influence on both the period and the amplitude of the Chandler oscillation and therefore causes a shift of the characteristic beat of free and forced oscillations. From spectral analyses of the time series by means of Fourier transformation, the relation between $\Re(k_2)$ and the Chandler period was obtained (Fig. 2, solid black line). Maximum agreement with the observed Chandler period of 434 days is reached for $\Re(k_2) = 0.3520$. Here, the correlation between the model time series for polar motion and C04 reaches 0.97. The relation between the Chandler period of the NCEP + ECCO forced model simulations and $\Re(k_2)$ is obviously non-linear.

In order to find the reason for the non-linear relationship, the experiment was repeated without any forcing, i.e. only the effect of rotational deformations was considered. Then DyMEG produces a damped oscillation with periods between 400 and 455 days, respectively. The resulting relation between $\Re(k_2)$ and the Chandler period (Fig. 2, dotted black line) is linear. Accordingly, the non-linearity in the NCEP + ECCO case is due to the interaction of free and forced polar motion. A similar relation between $\Re(k_2)$ and the Chandler period was derived from ECHAM3 + OMCT forced model runs. Neither in NCEP + ECCO nor in ECHAM3 + OMCT, there is increased excitation power in the Chandler frequency band. Nevertheless, there is a significant impact on the Chandler period.

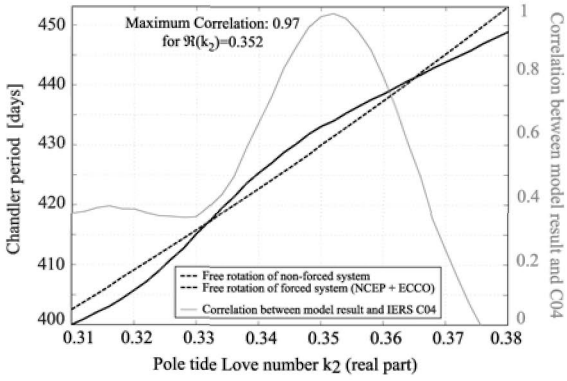


Fig. 2: Relation between $\Re(k_2)$ and the Chandler period as produced by DyMEG in case of forced and unforced conditions. Maximum correlation with the C04 series (scale on the right) is reached for $\Re(k_2) = 0.3520$.

A similar analysis was performed for the imaginary part of k_2. The value of the real part was set to 0.3520. The imaginary part $\Im(k_2)$ was increased equidistantly from 0 (no damping) to 0.0060. A small value of $\Im(k_2)$ causes a slow decrease of the Chandler amplitude, whereas a large value of $\Im(k_2)$ leads to a strong diminution of the Chandler wobble after few years. Unforced results of DyMEG feature damped oscillations, which have a constant period of 434 days but differ with respect to damping. In general, damping is expressed in terms of a so-called quality factor Q. From the damped oscillations of the unforced results, the factor Q was assessed by a least-squares fit (Seitz et al., 2004). For $\Im(k_2) = 0$, Q moves towards infinity, for $\Im(k_2) = 0.0060$ its value is 48 (Fig. 3).

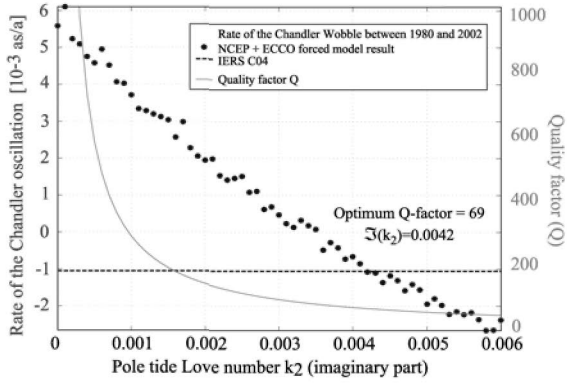


Fig. 3: Relation between $\Im(k_2)$ and the rate of the Chandler amplitude from 1980 until 2002. Optimum agreement between C04 (dotted) and the model result is achieved for $\Im(k_2) = 0.0042$ ($Q = 69$).

In order to assess the optimum damping for the NCEP + ECCO driven system, an annual oscillation with constant amplitude and a Chandler oscillation with a period of 434 days together with an amplitude rate were fit to the model result by means of least-squares adjustment. The resulting Chandler amplitude rates from each model run were compared with the respective value derived from C04 for the time 1980 - 2002 (Fig. 3). Optimum agreement was achieved for $\Im(k_2) = 0.0042$ which corresponds to a quality factor of $Q = 69$. The observed amplitude rate of the Chandler wobble between 1980 and 2002 is -1×10^{-3} as/a.

For ECHAM3 + OMCT, maximum agreement was reached for $\Im(k_2) = 0.0043$ ($Q = 68$) between 1975 and 1994. Hence, the power of the respective excitation series seems slightly higher in this case. But the difference between both model runs is marginal. Summing up the above results, the pole tide Love number k_2 which will be used in the further investigations with DyMEG is set to 0.3520 + 0.0042i.

3.2 Effect of the initial values

In DyMEG, the Liouville differential equation is solved by numerical integration. Hence, initial values $m_i(t = t_0)$ have to be provided. As mentioned above, the initial values are deduced from the C04 series of the IERS. In order to assess the effect of inaccurate initial values on the solution, each of the $m_i(t = t_0)$ was independently varied by uniformly distributed random numbers within $\pm 3\sigma_i$ around the respective C04 value. The standard deviations σ_i of each m_i were calculated from an interval of 30 days around the starting date $t = t_0$. Different starting dates were considered. The simulations were performed in half-yearly steps from 1973 until 1994 for ECHAM3 + OMCT and from 1980 until 2001 for NCEP + ECCO.

This investigation showed that the modification of the initial values within the $\pm 3\sigma_i$-interval for a single starting point does not have a large effect on the resulting time series. In general, the results of 30 model runs which were performed for each of the atmosphere-ocean combinations showed similar results. Deviations between different time series are maximum at the beginning of the simulations (due to the starting situation). But as convergence increases with time, DyMEG seems to reach a steady state.

However, the choice of the starting date seems to be more critical. Fig. 4 displays RMS values of the difference between the model result and C04 for polar motion (x-component) as well as the corresponding correlations against the varying starting date of the simulation. The model runs end at 31.12.1994 (ECHAM3 + OMCT) and 1.3.2002 (NCEP + ECCO).

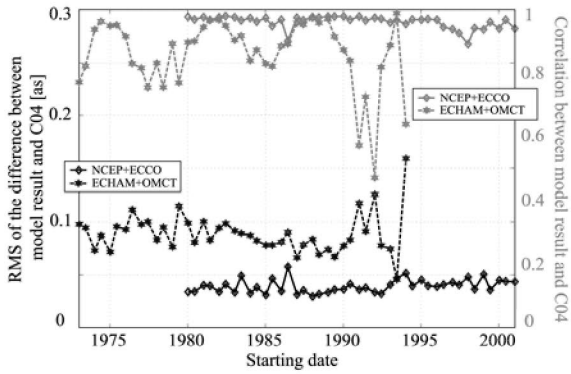


Fig. 4: RMS values of the difference between model results and C04 (black) and respective correlations (grey) for different starting dates.

In general, the NCEP + ECCO results are in better agreement with C04 than those of ECHAM3 + OMCT. Obviously, certain starting dates lead to better results than other ones. For ECCO + NCEP, a start at 1.6.1986 is disadvantageous whereas the neighbouring starting dates lead to good results. For ECHAM3 + OMCT the correlations show an oscillation, which is not visible in the RMS values. This oscillation does not show up in the case of NCEP + ECCO. Accordingly, the amplitudes of polar motion seem to be less affected than the phases and frequencies. During the last few years of the respective simulations, the correlations decrease because the considered parts of the polar motion series are rather short. Hence this effect seems to be an artefact.

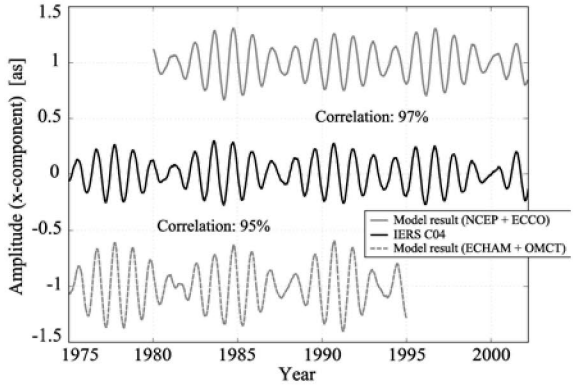


Fig. 5: Model results for DyMEG driven by NCEP + ECCO (top) and ECHAM3 + OMCT (bottom) in comparison with the geodetic observations C04 (middle). For better comparability, the model results are shifted by ±1as.

If the simulations are started at advantageous dates, the model results for polar motion are in good agreement with geodetic observations (Fig. 5). The correlations between the C04 values and the two displayed model runs are 0.95 and 0.97 respectively. As the annual signal of the ECHAM3 + OMCT result is too strong compared to the observations, the RMS of the difference to C04 is higher than for NCEP + ECCO (cf. Fig. 4).

3.3 Effect of variations in the initial tensor of inertia of the basic Earth ellipsoid

The influence of the choice of $\underline{\mathbf{I}}_0$ on the model results for polar motion was tested in two steps.

First, the model results for a triaxial ellipsoid of inertia ($A \neq B$) were contrasted with a simplified biaxial ($A = B$) solution. NCEP + ECCO forcing was applied in both cases. Although the difference between A and B is marginal, it was found that the Chandler period is shortened about two days if a rotationally symmetric ellipsoid of inertia $\underline{\mathbf{I}}_0$ is introduced into DyMEG. Vice versa, in order to lengthen the biaxial Chandler period to the value derived from the triaxial approach, the pole tide Love number has to be adapted (cf. 3.1). According to the results of this study, the appropriate value for $\Re(k_2)$ is 0.3550 in the case of a biaxial ellipsoid of inertia. The results are compiled in Table 1.

Table 1. Periods of the free Earth rotation as derived from DyMEG for a biaxial and triaxial ellipsoid of inertia $\underline{\mathbf{I}}_0$ in dependence of $\Re(k_2)$

	$\Re(k_2) = 0.3520$	$\Re(k_2) = 0.3550$
$A = B$	432	434
$A \neq B$	434	436

In a second step, the values of A, B, and C were changed as well as the directions of the principal axes of inertia with respect to the axes of the terrestrial reference system. Therefore, estimates of the Earth's tensor of inertia were introduced which are based on the recent gravity field solutions JGM-3, EGM96, GRIM5-S1, and GRIM5-S1CH1 (Marchenko and Schwintzer, 2003). In these solutions, the direction $\lambda(A)$ of the principal axis of inertia for the equatorial moment A varies between 345.0709° and 345.0712°. In order to study the sensitivity of DyMEG with respect to $\lambda(A)$, this angle was increased from 344.5° to 345.5° in 50 equidistant steps.

Neither the variation of A, B, and C nor the variation of $\lambda(A)$ within these reasonable limits led to significant changes of the resulting polar motion series. Therefore these parameters of the triaxial tensor of inertia $\underline{\mathbf{I}}_0$ are considered uncritical.

4. Conclusions

The sensitivity of DyMEG with respect to several input parameters as well as the corresponding reliability of the numerical results were assessed within this paper.

The results showed that period and damping of the Chandler wobble as derived from the model strongly depend on the pole tide Love number k_2. From various model runs, the value $k_2 = 0.3520 + 0.0042i$ was found to yield optimum agreement of the model result for polar motion with the C04 series of the IERS. While the value of the real part of k_2 is in good agreement with recent studies, the value of the imaginary part slightly differs from the one given in McCarthy (2003).

The sensitivity of the model Chandler period with respect to the real part of k_2 is very high. A change of this parameter by 1% results in a shift of the period of about two days. Hence, the inference from the Chandler period to $\Re(k_2)$ is relatively precise.

So far, only atmospheric and oceanic angular momentum variations were considered in DyMEG. Mass redistributions in other components of the Earth system like, e.g., land hydrology are still lacking. Hence, oscillations caused by these effects cannot be reproduced by the model. This might yield biases and thus affect the fitting procedure of the deformation Love number.

The variation of the triaxial initial tensor of inertia $\underline{\mathbf{I}}_0$ as well as the modification of the initial values $m_i(t = t_0)$ did not significantly change the results. However, the system is sensitive to the starting date. The correlation between the results from successive NCEP + ECCO driven model runs and C04 showed only slight variations whereas the ECHAM3 + OMCT forced results strongly depend on the starting date. The dependence of the quality of the resulting time series on the starting date will be subject to further investigations. When the simulations are started at advantageous dates the correlation between the model time series and the C04 series reaches 0.95 for ECHAM3 + OMCT and 0.97 for NCEP + ECCO (Fig. 5). As both resulting polar motion series show an undamped beat between free and forced polar motion, it can be concluded, that the atmospheric and oceanic excitation series are able to maintain the Chandler amplitude.

Acknowledgements

This paper was developed within a project funded by DFG grant DR 143/10. The authors thank M. Thomas (Technische Universität Dresden) and J. Stuck (Universität Bonn) for providing OMCT and ECHAM3 data sets. NCEP reanalysis data was provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA. Moreover we thank B. Richter (DGFI) for proof-reading and his helpful remarks.

References

Brzezinski A. (2001). Diurnal and subdiurnal terms of nutation: a simple theoretical model for a nonrigid Earth. In: Proceedings of the Journées Systèmes de Référence Spatio-temporels 2000, pp. 243-251, ed. Capitaine N., Paris.

Gross R. (1992). Correspondence between theory and observations of polar motion. Geophys. J. Int., 109, 162-170.

Kalnay E., Kanamitsu M., Kistler R. et al. (1996). The NCEP/NCAR 40-Year Reanalysis Project. Bull. Amer. Meteor. Soc., 77, 437-471.

Marchenko A., Schwintzer P. (2003). Estimation of the Earth's tensor of inertia from recent global gravity field solutions. J. Geodesy, 76, 495-509.

Mathews P., Herring T., Buffet B. (2002). Modelling of nutation and precession: New nutation series for nonrigid Earth and insights into the Earth's interior. J. Geophys. Res., 107, 10.1029/2001JB000390.

McCarthy D. (2003). IERS Conventions 2000. IERS Technical Note, Paris.

Moritz H., Mueller I. I. (1987). Earth Rotation. Ungar Publishing Company, New York.

Roeckner E., Arpe K., Bengtsson L. et al. (1992). Simulation of the present-day climate with the ECHAM model: Impact of the model physics and resolution. Tech. Rep. No. 93, Max-Planck-Institut für Meteorologie, Hamburg.

Sasao T., Okubo S., Saito M. (1980). A simple theory on the dynamical effects of a stratified fluid core upon nutational motion of the earth. In: Nutation and the Earth's Rotation, eds. Fedorov E., Smith M., Bender, P., IAU Symposia, 78, D. Reidel, Kiev, 165-183.

Seitz F., Kutterer H. (2002). Numerical Solutions for the Non-Linear Liouville Equation. In: Vistas for Geodesy in the New Millennium, eds. Adam J., Schwarz K.-P., IAG Symposia, 125, Springer, Berlin, 463-468.

Seitz F., Stuck J., Thomas M. (2004). Consistent Atmospheric and Oceanic Excitation of the Earth's Free Polar Motion. To be published in Geophys. J. Int.

Smith M., Dahlen F. (1981). The period and Q of the Chandler wobble. Geophys. J. R. astr. Soc., 64, 223-281.

Stammer D., Wunsch C., Giering R. et al. (2003). Volume, heat and freshwater transports of the global ocean circulation 1993-2000, estimated from a general circulation model constrained by World Ocean Circulation Experiment (WOCE) data. J. Geophys. Res., 108, 10.1029/2001JC001115.

Thomas M., Sündermann J., Maier-Reimer E. (2001). Consideration of ocean tides in an OGCM and its impacts on subseasonal to decadal polar motion excitation. Geophys. Res. Lett., 28, No. 12, 2557-2560.

Filters to estimate water storage variations from GRACE

K.-W. Seo[1], C. R. Wilson[12], J. Chen[2], J. S. Famiglietti[3], M. Rodell[4]

[1] Department of Geological Sciences, University of Texas at Austin ,e-mail: kiweon@geo.utexas.edu
[2] Center for Space Research, University of Texas at Austin
[3] Department of Earth System Sciences, University of California at Irvine
[4] Goddard Space Flight Center, NASA

Abstract. We examine the problem of obtaining average surface mass load changes in a local area (for example, water content in a river basin) from time-variable Stokes coefficients available from the GRACE mission. A basin function is unity inside a defined geographical region, and zero outside, and can be represented exactly only if Stokes Coefficients of all degrees and orders are available. GRACE Stokes coefficients will contain errors that generally increase with degree, and will be of limited degree range, perhaps to 100 or so. Load variations within a basin should be estimated by minimizing some quantity that accounts for both GRACE measurement error, and leakage error, associated with a finite degree range. To understand this problem, we use Fourier series, the 1-D equivalent to spherical harmonics. The solution of this problem is not unique, and we examine several different approaches. We use Monte Carlo experiments to test the performance of various methods derived. Time series of gridded soil moisture, snow, ocean bottom pressure, atmospheric surface pressure and Gaussian random noise are utilized in an experiment to recover load variations within the Nile basin.

Key words. GRACE – surface mass load change – spatial filter – basin function – leakage error

1 Introduction

Variations of GRACE's Stokes coefficients represent global gravity changes due to redistribution of air and water mass, and other sources. We assume that the effect of atmospheric surface pressure is eliminated, but imperfectly, leaving some residual that is characterized as noise. We ignore contributions to Stokes Coefficient changes other than air and water. After correcting for atmospheric mass effects, the residual Stokes coefficients can be used to estimate water mass fluctuations within a defined region, called a basin. The exact basin function is 1 inside and 0 outside and can be represented by spherical harmonics. The effective basin function is a spatial filter that approximates the exact, and is used to extract mass load changes in a basin, considering the effects of errors and a limited range of spherical harmonics. The estimation of an effective basin function has been addressed previously by Famiglietti *et al.* (2001) and Swenson and Wahr (2002). We examine the problem because it does not have a unique solution, implying that other useful effective basin functions may exist that are best in the sense of minimizing some cost function. The problem involving 2-D spherical harmonics is analogous to the 1-D problem of Fourier series. Fourier series are easier to use in developing new methods because of their simplicity. In this study, several approaches to estimate basin functions are investigated with Fourier series, and then expanded to 2-D spherical harmonics where the performance of the basin functions are tested.

2 Fourier series analysis of basin function estimates

Assume that GRACE products are c_l and s_l, Fourier series coefficients with the usual properties

$$L(x) = \sum_{l=0}^{N} [c_l cos(lx) + s_l sin(lx)] \tag{1}$$

where $L(x)$ is surface mass load changes and cos and sin are normalized

$$\int_0^{2\pi} \begin{bmatrix} cos(lx) \\ sin(lx) \end{bmatrix} \begin{bmatrix} cos(mx) \\ sin(mx) \end{bmatrix} dx = \begin{pmatrix} 0, l \neq m \\ 1, l = m \end{pmatrix} \tag{2}$$

The basin function is defined by a Fourier series

$$B(x) = \sum_{l=0}^{N} [a_l cos(lx) + b_l sin(lx)] \tag{3}$$

The basin function is exact when an infinite number of coefficients are used. However, high order coefficients for a_l and b_l can not be fully utilized because GRACE measurement error has a blue spectrum, and coefficients are limited to some finite order N. Figure (1) shows an exact basin function and a truncated basin function. The familiar Gibb's effect

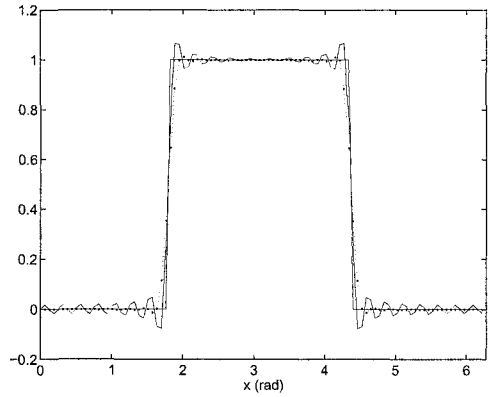


Fig. 1. The truncated basin function $B(x)$ with $N = 25$ oscillates with respect to the exact basin function. After *sinc* function modification to the truncated basin function (dotted line), the ripples are smoothed.

appears inside and outside the basin. The mean surface mass load change over the basin is

$$\bar{L} = \int_0^{2\pi} L(x)B(x)dx = \sum_{l=0}^{N}[c_l a_l + s_l b_l] \tag{4}$$

The measurement noise in c_l and s_l and the ripples of the basin function of a_l and b_l will corrupt the estimate of the load within the basin.

2.1 Suppressing ripples

Ripples due to a finite range of Fourier series coefficients (the Gibbs effect) can be smoothed by a moving average of the truncated basin function, equivalent to multiplication of Fourier coefficients by a sinc function (Hamming (1962)).

$$\hat{B}(x) = \frac{1}{2\pi/N}\int_{x-\pi/N}^{x+\pi/N} B(x')dx'$$

$$= \sum_{l=0}^{N} sinc(\frac{l}{N})[a_l cos(lx) + b_l sin(lx)] \tag{5}$$

The dotted line in figure (1) is the basin function adjusted in this way, showing that the oscillations are reduced. Since the amplitude of a sinc function decreases with l, the contribution of higher l coefficients, where measurement error is larger, is also reduced.

2.2 Estimates with noisy data (Swenson & Wahr method)

2.2.1 Estimates with known load variance and autocorrelation Let $A(x)$ be the normalized (unity at zero-lag) spatial autocorrelation function of the load, with a Fourier series representation given by G_l. A cost function minimized to determine an effective basin function coefficients, $\hat{a}_l$ and $\hat{b}_l$ is

$$\Phi = V_0 \sum_{l=0}^{N} G_l[(a_l - \hat{a}_l)^2 + (b_l - \hat{b}_l)^2] + \sum_{l=0}^{N} E_l^2(\hat{a}_l^2 + \hat{b}_l^2) \tag{6}$$

Here E_l is the measurement error spectrum and V_0 is the zero-lag autocorrelation of the load, that is, its variance. The solution minimizing this cost function is found by taking the deriviative of the least-squares cost function and setting it to zero, leading to

$$\begin{pmatrix}\hat{a}_l \\ \hat{b}_l\end{pmatrix} = \frac{G_l}{G_l + E_l^2/V_0}\begin{pmatrix}a_l \\ b_l\end{pmatrix} \tag{7}$$

When there is no noise these estimates are exactly the same as obtained with the exact basin coefficients, truncated to finite order N, as in (4).

2.2.2 Estimates employing a Lagrange Multiplier An alternative cost function is

$$\Phi = \sum_{l=0}^{N}[(a_l - \hat{a}_l)^2 + (b_l - \hat{b}_l)^2] + \lambda[\sum_{l=0}^{N} E_l^2(\hat{a}_l^2 + \hat{b}_l^2) - \Delta^2] \tag{8}$$

where Δ is a desired level of measurement error over the basin, and λ is a Lagrange multiplier. The solution minimizing this cost function is

$$\sum_{l=0}^{N} E_l^2 \frac{a_l^2 + b_l^2}{(1 + \lambda E_l^2)^2} = \Delta^2 \tag{9}$$

$$\begin{pmatrix}\hat{a}_l \\ \hat{b}_l\end{pmatrix} = \frac{1}{1 + \lambda E_l^2}\begin{pmatrix}a_l \\ b_l\end{pmatrix} \tag{10}$$

λ needs to be obtained in (9) by iterative solution before a solution can be obtained in (10) because equation (9) is implicit.

2.3 Estimates using climate model data

Let the diagonal matrices $\tilde{S}$ and $\tilde{E}$ represent signal and error values as a function of Fourier order l along their diagonals. Climate models might provide $\tilde{S}$. m is the vector of coefficients of the truncated exact basin function and $\hat{m}$ is a vector to be estimated, consisting of coefficients of the effective basin function. The cost function including signal leakage and measurement error is

$$\Phi = (\tilde{S}m + H - \tilde{S}\hat{m})^T(\tilde{S}m + H - \tilde{S}\hat{m}) + (\tilde{E}\hat{m})^T(\tilde{E}\hat{m}) \tag{11}$$

Here, H represents loads in a basin exceeding the maximum order (wavenumber) of $\tilde{S}$ and m. In this cost function, the first term is leakage error and the second term is measurement error. H is unknown but may be relatively small compared to the other parts of the cost function and is ignored. The Gibb's effect will arise from the finite number of coefficients, and to reduce Gibbs oscillations, the difference operator ∂ given by Constable *et al.* (1987)

$$\tilde{\delta} = \begin{vmatrix} 0 & 0 & 0 & \dots & \dots & 0 \\ -1 & 1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 1 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & -1 & 1 \end{vmatrix} \quad (12)$$

is used to modify the cost function by providing a measure of solution roughness.

Let $\tilde{G}$ be a matrix of sinusoidal functions sampled at each x and l, $\tilde{\delta}\tilde{G}\hat{m}$ be the roughness error due to the Gibb's effect. The new cost function, including penalties of signal leakage, measurement error and Gibb's effect roughness is

$$\Phi = (\tilde{S}m - \tilde{S}\hat{m})^T(\tilde{S}m - \tilde{S}\hat{m}) + (\tilde{E}\hat{m})^T(\tilde{E}\hat{m}) + \lambda(\tilde{\delta}\tilde{G}\hat{m})^T(\tilde{\delta}\tilde{G}\hat{m}) \quad (13)$$

where λ is a scale factor for the roughness error. Taking the derivative with respect to elements of $\hat{m}$ to find minimum value of (13), the solution for $\hat{m}$ is

$$(\tilde{W}_s + \tilde{W}_e + \lambda\tilde{D})\hat{m} = \tilde{W}_s m \quad (14)$$

in which $\tilde{W}_s = \tilde{S}^T\tilde{S}$, $\tilde{W}_e = \tilde{E}^T\tilde{E}$ and $\tilde{D} = (\tilde{\delta}\tilde{G})^T(\tilde{\delta}\tilde{G})$.

3 A test using spherical harmonics

The spherical harmonic expansion of surface mass load is (Swenson and Wahr (2002))

$$\bar{L} = \frac{4\pi R\rho_E}{3\Omega_B}\sum_{l=0}^{N}\sum_{m=0}^{l}\frac{2l+1}{1+k_l}(\Delta C_{lm}B^c_{lm} + \Delta S_{lm}B^s_{lm}) \quad (15)$$

where ΔC_{lm} and ΔS_{lm} are Stokes coefficients generated from observations, and containing error, B^c_{lm} and B^s_{lm} are effective basin coefficients estimated by one of the methods described in the previous section, k_l is load Love number describing load deformation, R and ρ_E are mean radius and density of Earth, and Ω_B is the solid angle of the basin. Geoid variation (ΔN) and basin function (B) in a spherical harmonic expansion from (15) are

$$\Delta N(\theta,\phi) = R\sum_{l=0}^{N}\sum_{m=0}^{l}\tilde{P}_{lm}(cos\theta)(\Delta C_{lm}cosm\phi + \Delta S_{lm}sinm\phi) \quad (16)$$

$$B(\theta,\phi) = \sum_{l=0}^{N}\sum_{m=0}^{l}\tilde{P}_{lm}(cos\theta)(B^c_{lm}cosm\phi + B^s_{lm}sinm\phi) \quad (17)$$

To simulate ΔC_{lm} and ΔS_{lm}, soil moisture, snow and ocean bottom pressure data and model predictions for the period 1981 to 1992 are used. The GLDAS (Global Land Data Assimilation Scheme) is used to represent soil moisture and snow, and the ECCO (Estimating the Circulation and Climate of the Ocean) is used to estimate ocean bottom pressure. Simulated Stokes coefficients are corrupted using Gaussian zero-mean random numbers. Error associated with removing atmospheric surface pressure is estimated as the difference between two comparable numerical model estimates (Wahr *et al.* (1998)), NCEP (National Centers for Environmental Prediction) and GEOS (Goddard EOS assimilation system).

The spatial spectral properties of a time series of gridded data can be described by the spectrum as a function of spherical harmonic degree, by including the contributions from all orders at each degree. The amplitude spectrum is denoted by σ_l. The spectrum of the annual variation depends on ΔC^a_{lm} and ΔS^a_{lm}, amplitudes of annual sinusoidal coefficients of Stokes coefficients, is

$$\sigma^a_l = R\sqrt{\sum_{m=0}^{l}(\Delta C^{a\,2}_{lm} + \Delta S^{a\,2}_{lm})} \quad (18)$$

Figure (2) shows separate spherical harmonic degree spectra of annual variations in continental water storage, oceanic bottom pressure, atmospheric surface pressure, atmospheric surface pressure error, and the assumed GRACE measurement error. The spectra are similar to previous results by Wahr *et al.* (1998).

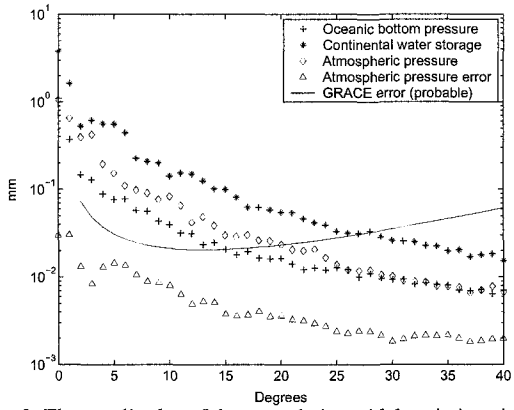


Fig. 2. The amplitudes of the annual sinusoidal variations in degree amplitudes of input signals and error. The unit is geoid.

GRACE products are expected to be monthly Stokes coefficient time series. (Dickey *et al.* (1999)). The root mean squared (RMS) of monthly time series of ΔC_{lm} and ΔS_{lm} give a spectrum as a function of spherical harmonic degree defined by

$$\sigma^t_l = R\sqrt{\sum_{m=0}^{l}\frac{1}{M}\sum_{t=1}^{M}(\Delta C^2_{lmt} + \Delta S^2_{lmt})} \quad (19)$$

where M is the total number of monthly values in the time series. Figure (3) compares the degree amplitude spectra of RMS Stokes coefficients with those from figure (2). The difference between the two spectra is due to the relative size of

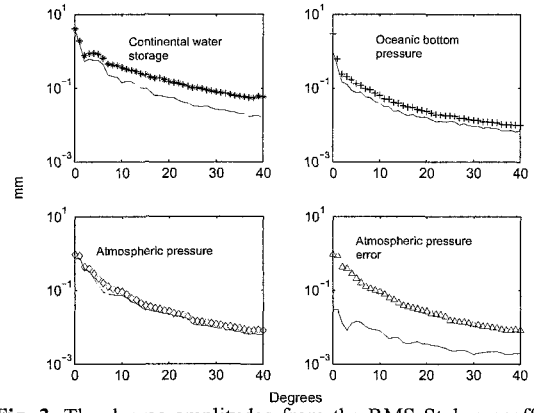


Fig. 3. The degree amplitudes from the RMS Stokes coefficients. The legends are same as in figure (2). The solid line at each section is the curve from figure (2).

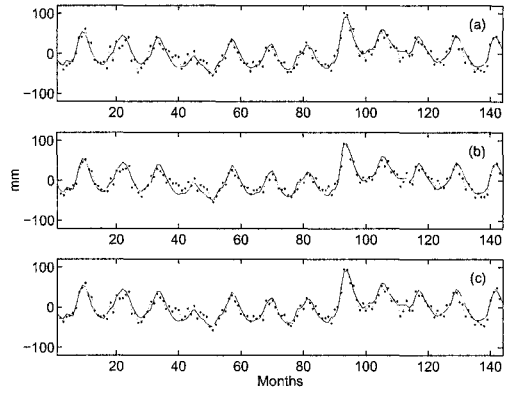


Fig. 4. The water thickness variations in Nile basin. The solid lines are true values and dotted lines are recovered values from the basin functions, B1, B2 and B3.

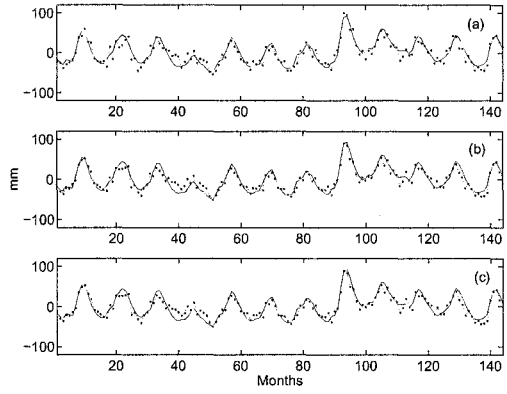


Fig. 5. The water thickness variations in Nile basin. The solid lines true values and dotted lines are recovered values from the basin functions, B4, B5 and B6.

annual variations to those at all other temporal frequencies. That is, if the two curves in Figure 3 are about the same size, then the variations are dominantly annual. For atmospheric surface pressure and oceanic bottom pressure, Figure 3 shows amplitudes are similar at most spherical harmonic degrees. However, they differ greatly at zero degree in the case of ocean bottom pressure, probably due to mass conservation issues in the ECCO model. In any case, because the curves are similar at most spherical harmonic degreees, the ocean and atmosphere gravity signals are dominated by annual changes. Figure 3 shows that for continental water storage, annual effects appear to be less dominant, especially with increasing order, that is, at smaller spatial scale. Another important result from Figure 3 is that RMS atmospheric pressure error is significantly bigger than the spectrum for the annual component of error. This indicates that atmospheric pressure will be a serious problem at higher temporal frequencies (shorter periods), when estimating water storage changes in a basin.

4 Results

The water mass changes over the Nile river basin are recovered with respect to 6 different spatial filters described in section 2. These 6 are : B1, the truncated basin function; B2, the basin function with sinc function modification in (2.1); B3 and B4, the basin functions estimated by Swenson and Wahr (2002) in (2.2.1) and (2.2.2) respectively; B5 and B6, the basin functions in (2.3) without and with smooth operators respectively.

V_0 for the basin function B3 is the variance of the time series of mean values in the basin. G_l is assumed to be a Gaussian function varying from unity at center to $1/2$ value at 4000 km radius. In future studies, we expect to use actual climate model data to set these parameters. For B4, the desired error, Δ, is set at the sum of measurement and surface pressure errors using the truncated basin function. For B6, it is necessary to define λ, which gives relative weight to roughness error. The units of the error sources are different, so the value here is $\lambda = 10^{-24}$. In practice we used trial and error to select λ. In the future a more systematic method will be used to select λ. Figures (4) and (5) show variations of true and recovered water thickness variations over the Nile basin. Each time series of recovered water thickness change fits well with the true line.

The performances of the 6 basin functions are examined using four different measures (total error:e_T, leakage error:e_L, measurement error:e_M and atmospheric error:e_A). The estimated variation of water load over the basin, $\bar{L}$, is

$$\bar{L} = L_c + L_o + L_m^E + L_a^E \tag{20}$$

where L_c and L_o are from continental water storage and oceanic bottom pressure, and L_m^E and L_a^E are GRACE measurement error and atmospheric surface pressure error. L_o is entirely leakage of ocean load into the estimate from sidelobes of the effective basin function. L_c is the sum of the true mean load T in the basin plus leakage from outside the basin and irregular sampling in the basin due to sidelobes. For example,

$$e_M = \frac{E[\mid L_m^E \mid]}{E[\mid T \mid]} \tag{21}$$

$$e_L = \frac{E[\mid L_c + L_o - T \mid]}{E[\mid T \mid]} \tag{22}$$

Table 1. The error over Nile basin

	B1	B2	B3	B4	B5	B6
e_T	0.4213	0.3759	0.3913	0.4107	0.3693	0.3669
e_L	0.0350	0.1043	0.0390	0.0373	0.1002	0.1150
e_M	0.3323	0.2517	0.2923	0.3186	0.2388	0.2304
e_A	0.2337	0.2276	0.2324	0.2311	0.2208	0.2195

Since the errors are related to water volume, an absolute mean measure seems more appopriate than a variance measure. Table (1) summarizes the errors. The basin function B1 provides the least leakage error despite the oscillations of the basin function. This is probably because water storage variations inside the Nile basin are small because the basin is arid, as are adjacent basins. Nearby ocean effects are probably also small. Thus all 6 basin functions behave similarly, and difference in performance are only evident from the numbers in table (1).

Atmospheric error is comparable to measurement error, but it is hard to design an excellent spatial filter because the two errors have spherical harmonic spectra with opposite slopes over small ranges (measurement error is blue and atmospheric surface pressure error is red).

5 Discussion

GRACE measurement error and atmospheric surface pressure error appear to be the most significant in this study. However, another error source, alias error may also seriously corrupt GRACE- Stokes coefficients. The alias problem needs to be considered in a future study. The GRACE alias error will arise from high frequency mass load variations. With monthly values, signal variations at periods less than two months will likely be aliased into longer period signals, and there will be alias problems between various spherical harmonic degrees and orders. Once the nature of the aliasing error spectrum is known, cost functions investigated in this paper may be used to estimate the basin function by considering the alias problem. For example, the smoothing operator (2.3) applied to longitudinal and latitudinal directions independently might be useful to reduce alias error effectively.

Acknowledgements. This research was funded by NASA grants NAG5-8798 and NGT5-30447. Additional support was provided by the Geology Foundation at the University of Texas.

References

Constable SC, Parker RL and Constable CG (1987) Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, Geophysics, 52(3), 289-300

Dickey JO, Bentley CR, Carton JA, Eanes J, Herring TA, Kaula WM, Lagerloef SE, Rojstaczer S, Smith WHF, Van Den Dool HM, Wahr JM, Zuber MT (1999) Gravity and the hydrosphere: new frontier, Hydrological Sciences, 44(3), 407-415.

Famiglietti JS, Seo KW, Wilson CR and Rodell M (2001) Remote sensing of terrestrial water storage variations using GRACE (abstract), EOS Trans. AGU, 82(47). Fall Meet. Suppl., F441

Hamming RW (1962) Numerical methods for scientists and engineers, McGraw-Hill, New York

Swenson SC and Wahr J (2002) Methods for inferring regional surface-mass anomalies from satellite measurements of time-varying gravity, J. Geophy. Res., 107(B9)

Wahr J, Molenaar M. and Bryan F. (1998) Time variability of the Earth's gravity field: Hydrological and oceanic and their possible detection using GRACE, J. Geophys. Res., 103(B12), 30,2005-20,230, 1998

Author Index

Abd-Elmotaal H.A. 374, 471
Abdel-Salam M. 77
Abidin H.Z. 561
Ádám J. 214
Akyilmaz O. 582
Alves Costa S.M. 32
Amagai T. 201
Amos M.J. 338
Andersen O.B. 326
Andreas H. 561
Andritsanos V.D. 332
Angermann D. 3, 9, 576
Aoki H. 567
Arabelos D.N. 310
Araki H. 163
Asari K. 152, 163
Awange J.L. 524

Barnes J. 83
Beavan J. 38
Beutler G. 169, 175
Bilker M. 15
Bingley R.M. 26
Blankenberg L.E. 71
Blewitt G. 518
Blick G.H. 38, 338
Blitzkow D. 374
Boehm J. 125, 131
Bouman J. 265
Brzeziński A. 591

Chen J. 607
Chen K. 77
Chmelina K. 105
Claessens S.J. 459
Clarke P. 518
Cosser E. 83, 89
Crook C. 38

Deng X. 314
Denker H. 214, 248, 368
Dettmering D. 60
Diament M. 446
Dietrich R. 195
Djaja R. 561
Dodson A.H. 26, 83, 89
Drewes H. 3, 32, 544
Dudevoir K.A. 119

Ellmann A. 403, 409
El-Sheimy N. 506
Engelhardt H.G. 125
Fabert O. 362

Fairhead J.D. 374
Famiglietti J.S. 607
Fashir H.H. 374
Featherstone W.E. 300, 314, 338, 350, 459
Feissel-Vernier M. 207
Felus Y.A. 417
Feuchtinger M. 189
Földváry L. 214, 288
Font G. 398
Forsberg R. 306
Fotopoulos G. 440
Frey S. 137
Fukuda Y. 223, 236, 524
Fukushima T. 141
Fukuzaki Y. 146
Funazaki K. 163

Gamal M. 561
Gao Y. 77
Gebhard H. 60
Gerasimenko M.D. 538
Gerlach C. 288
Gerstl M. 3
Goto Y. 44
Gouda N. 163
Grafarend E.W. 524
Grant D. 38
Grebenitcharsky R.S. 320
Gruber T. 288
Gubanov V. 125

Haagmans R. 265
Haas R. 125
Haberler M. 95
Hanada H. 152, 157, 163
Heck B. 294
Heidbach O. 544
Heki K. 163
Higashi T. 223, 236
Hirose K. 561
Hobiger T. 131
Holota P. 452
Hugentobler U. 169, 175

Ichikawa R. 113
Ilk K.H. 189, 494
Imanishi Y. 223
Ishihara T. 229
Ito T. 229
Iwata T. 152, 157, 163
Izumi T. 157

Jäggi A. 169, 175
Jamet O. 446
Janssen V. 533
Jensen A.B.O. 65
John S. 374
Jokela J. 15
Joseph J.E. 229

Kahmen H. 101, 105
Kanamori H. 163
Kaniuth K. 32
Kasahara M. 538
Kaschenz J. 500
Kawano N. 152, 157, 163
Kawawa H. 146
Kearsley A.H.W. 259
Kelm R. 3
Khan S.A. 556
Kikuchi F. 157
Kjorsvik N.S. 71
Knudsen P. 326, 483, 556
Kobayashi S. 236
Koivula H. 15
Kokubo K. 223
Kolaczek B. 591, 597
Komazawa M. 229, 236
Komuro H. 236
Kondo T. 113, 119
Kono Y. 152, 157, 163
Koop R. 265
Kotsakis C. 434, 440
Koyama Y. 113, 119
Krügel M. 3, 9
Krynski J. 48
Kudo T. 236
Kudryavtsev S.M. 465
Kuhn M. 350, 386
Kulkarni M.N. 137
Kumar V.K. 229
Kurihara S. 146
Kusumoto S. 229
Kutterer H. 54, 582, 601

Lanotte R. 125
Lavallée D. 518
Lee H-K. 83
Lehmann A. 195
Lemoine F. 207
Liebsch G. 195
Liu Q. 392
Löcher A. 494
Luo J. 283

MacMillan D. 125
Maruyama Y. 561
Matsumoto K. 163
Matsuzaka S. 146, 201
Mautz R. 500
Mayer-Gürr T. 189, 254
Meisel B. 3, 9
Meng X. 83, 89
Merry C.L. 374
Miyagawa K. 146
Miyamachi H. 236
Miyazaki T. 163
Mueller F. 254
Müller H. 9
Müller J. 265
Murata Y. 236
Murdohardono D. 561

Nakada M. 236
Nakagawa I. 236
Nakayama E. 229
Namiki N. 152
Nastula J. 591, 597
Nawa K. 236
Negusini M. 125
Niemeier W. 101
Ning J.Sh. 283
Nishikori M. 44
Noda H. 152, 163
Noll C. 207
Noureldin A. 506
Novotny K. 195
Nozaki K. 567
Nurutdinov K. 518

Ohno I. 236
Oikawa J. 236
Okubo S. 488
Ollikainen M. 15
Omang O.C.D. 306
Osaki H. 113, 119
Oshita K. 567
Osman A. 506
Ovstedal O. 71

Peters T. 288
Pillai P. 137
Podmore F. 374
Ponte R.M. 591
Poutanen M. 15

Quintas M.C.L. 477

Rajiyowiryono H. 561
Reda Taha M.M. 506
Ries J. 207
Rizos C. 83, 533
Roberts G.W. 83, 89
Rodell M. 607
Roland M. 248, 368
Rothacher M. 181, 576
Rózsa Sz. 344
Rummel R. 288

Salstein D. 597
Sasaki T. 157
Satomura M. 44
Schaer S. 169
Schaffrin B. 417, 500
Schmid R. 576
Schmidt M. 362
Schön S. 54
Schuh H. 125, 131
Segawa J. 229
Seitz F. 601
Seitz K. 386
Sekido M. 113, 141
Sekizaki S. 229
Seo K.-W. 607
Shestakov N.V. 538
Shichi R. 236
Shimada S. 21, 44
Shum C.K. 362
Sideris M.G. 320, 380, 392, 398, 440
Sneeuw N. 265, 288
Soudarin L. 207
Souto Fortes L.P. 32
Sproule D.M. 259
Stary B. 356
Strykowski G. 306
Subarya C. 561
Sugihara M. 571
Sun W. 488, 512
Švehla D. 181, 288
Svendsen J.G.G. 71

Takashima K. 146
Takemoto S. 524
Tanabe T. 146
Tanaka T. 567
Tanaka Y. 201
Tavernier G. 207
Tazawa S. 163
Teferle F.N. 26
Tesmer V. 9, 576
Teunissen P.J.G. 422, 428
Thaller D. 576
Tocho C. 380, 398
Tomasi P. 125
Tóth Gy. 214, 344
Trimmer R. 326
Tscherning C.C. 265, 277, 310, 556
Tsoulis D. 356
Tsubokawa T. 163
Tsuji H. 146, 201
Tsuruta S. 163
Tziavos I.N. 214, 332, 434

Vereshchagina I. 125
Vergos G.S. 332, 380
Verhagen S. 428
Virtanen H. 15
Visser P.N.A.M. 265, 271
Völksen C. 32

Waithaka H.E. 538
Wang H. 550
Watanabe H. 236
Waugh A.I. 26
Weber G. 60
Whitney A.R. 119
Wild F. 294
Willis P. 207
Wilson C.R. 607
Wojciechowski A. 77

Xu C. 550
Xu J. 550

Yamada A. 201
Yamamoto A. 236, 242
Yamamoto Z. 152
Yano T. 163
Yokoyama T. 163
Yusa Y. 236

Zanimonskiy Y.M. 48

Keyword Index

Absolute gravimeter 223, 571
Accelerometer 283
Active fault 567
Active fault block 550
Affine coordinate transformation 95
Africa 374
Airborne gravimetry 254
Airborne sensors 71
Alert systems 101
Altimetric gravity anomalies 248
Altimetry 380
Altimetry-gravimetry boundary value problems 320
Analysis of time series 9
ANFIS 582
APRGP 201
Arctic Ocean 326
Argentina 398
Asia Pacific Regional Geodetic Project 201
Asymptotic solution 488
Atmospheric excitation 591
Atmospheric models 597

Baltic Sea 195
Basin function 607
Bathymetry 259
Best integer equivariant estimation 422
Best integer equivariant unbiased estimator 428
Biharmonic spline wavelets 500
British Isles 26
Butterworth filter 271

Calibration 223
CHAMP 189, 283, 288, 300
China continent 550
Clebsch-Gordan 518
Coastal waveform retracking system 314
Collocated sites 125
Collocation 544
Combination strategies 3
Computation efficiency 392
CONT02 125, 576
Continuous surface representation 500
Convolution 374
Co-seismic displacement 488
Critical parameters 601
Crossover adjustment 248

Damage detection 506
Datum 38
Dedicated satellite gravimetry 300
Deficiencies 3
Deformation 38
Deformation analysis 105
Deformation monitoring 533
Delta VLBI 113
Dense-network campaign 21
Densification 524
Density anomaly 471
Density distribution 446
Differential GPS 60
Discretization scheme 512
Dislocation theory 488
DORIS 9, 207
Downward continuation 254, 512
Drift 223

Earth rotation 582, 601
Earthquake 488
Earth's gravity field 452
Earth's mass-density distribution 350
EIGEN-2S 306
Ellipsoidal Boundary Value Problem 459
Ellipsoidal correction 409
Energy balance relations 189, 494
EOP 119, 146
Eötvös' torsion balance 214
Errors-in-variables 417
Eurasian Plate 26
e-VLBI 119

Fast spherical collocation 310
Fault mapping 242
FG5 571
FG5L 571
Finite distance 141
Finite Element (FE) model 89
Finite element method 544
Fixed constraints 417
Float and fixed estimator 428
Forward modelling 350
Frame transformation 518
Functional-analytic estimates 452
Fuzzy inference systems 582
Fuzzy system 95

G_i 374
GEDEX 137
Geodesy 137
Geodetic boundary-value problems 452
Geodynamic monitoring GPS network 538
Geoid 300, 320, 374, 398, 409, 512
Geopotential model 310, 459
Geothermal reservoir 571
GIS 477
Glacial Isostatic Adjustment 556
Global Positioning System (GPS) 21, 48

Global search algorithm 500
GNSS 71
GNSS ambiguity resolution 422, 428
GOCE 214
GOCE mission 310
GPS 9, 26, 32, 44, 54, 77, 89, 131, 146, 556
GPS augmentation 83
GPS positioning 65
GPS survey 561
GPS time series 15
GPS velocity field 544
GPS-meteorology 524
GRACE 189, 607
Gravimetric geoid 338
Gravity 157, 320, 326, 512
Gravity anomaly 236, 242
Gravity database 236
Gravity disturbance 229
Gravity field modeling 189
Gravity field recovery 356
Gravity field recovery performance 271
Gravity field validation 494
Gravity monitoring 571
Gravity prediction 259
Groebner basis 524
Gyroscopic model 601

Harmonic development 465
Helicopter gravimetry 229
Helmert's condensation models 294
Hokkaido 242
Horizontal Crustal Motions 26
Hydrological cycle 518

IDS 207
IGGOS 207
Imprecision 54
Improved sea surface height 314
Inland lakes 326
InSAR 561
International DORIS Service 207
Internet 60
Internet VLBI 113
Interval mathematics 54
Inverse gravimetric problem 446
Inverse problems 471
Ionosphere 131
Ionospheric corrections 533
IOST 332
Ishikari Lowland 242
Ishikari Plain 242
Isomorphic 518
Isostasy 356, 471
Isostatic modeling 477
ITRF 3, 9, 146, 201

Jacobi integral 189, 288, 494
Jakarta 561
Japan 242

Kernel compression 392
Kinematic orbit 181
Knowledge-based system 105
Kyushu district 236

Land subsidence 561
Landslide monitoring 95
Landslides 101
Land-to-sea consistency of gravity 229
Leakage error 607
Least squares modification 409
Least-squares collocation 332
LEO 181
Leveling 561
Liouville differential equation 601
Local ties 576
Long-term gravity changes 223
Long-wavelength gravity anomaly errors 300
Love number 518
Low Earth orbiter 169, 175
Lunar core 163

Marine geoid 380
Marine geoid modeling 332
Microgravity 567
MIMOST 380
Mixed-mode GPS 533
Modification 403
Molodensky 374
Moon 157
Mountain area 44
Multipolynomial resultant 524
Multiquadrics 500
Multi-resolution analysis 362
Multi-scale analysis 368

Natural hazards 101
Navigation 60
Netwon's integral 386
New Geoid models 306
New Zealand 38, 338
Non-linear programming task 538
North Atlantic region 306
Numerical integration 386
Numerical weather predictions (NWP) 65

Oceanic angular momentum 591
Oceanic excitation 591
Oceanographic model 195
Optimal design 538
Orbit determination 169
Orbit determination validation 494

Parameter estimation methods 362
Paraná basin 477
PCGIAP 201
Periodic variations 15
Photographic zenith tube (PZT) 163
Physical librations 163

POD 181
Poisson integral 512
Polar motion 591
Polar motion excitation 597
Polynomial ring 524
Positioning 60
Positioning accuracy 48
Precise point positioning 77
Prediction 582
Pseudolite 83

Quasi-geoid 374

RARR 157
Real-time 60
Recent crustal movements 544
Recommendations 3
Reduced-dynamic orbit determination 175
Reference Earth model 356
Regional gravity field recovery 189
Regularization 403
Response function 471
RSO 283
RTK 77

Sapporo 242
Satellite altimetry 195, 320, 326
Satellite gravity gradiometry 294
Satellite radar altimeter waveforms 314
Satellite-to-Satellite-tracking 189
Sea floor seismic fault 229
Sea Surface Heights 380
SELENE 157
Selenodesy 157
Sensitivity analysis 601
Shipborne gravity 380
Shipborne gravity observations 248
Single-step approach 169
Singularity 392
SIRGAS 544
SLR 9
Soft computing 582
South America 32, 544
Southwest Japan 236
Space clock 181
Space VLBI 137
Space-borne gravity gradiometer 271
Spacecraft navigation 141
Spatial filter 607
Spectral analysis 48
Spectral method 465
Spherical earth 488
Spherical harmonics 362, 386, 459
Spherical wavelets 362, 368
Statistical analysis 48
Stochastic orbit parametrization 175
Stokes 368, 392
Stokes' formula 403, 409
Structural deformation 83
Structural deformation monitoring 89
Structural health monitoring 506
Sub-daily EOPs 576
Superconducting gravimeter 223
Surface mass load change 607
Synthetic Earth gravity model 350
Systematic errors 54

TEC 131
Tectonic strain & stress field 550
Terrain corrections 392
Terrain reductions 398
Terrestrial Reference Frame 9, 32
Tesseroid 386
Theory of Relativity 141
Tide gauge 195
Tide generating potential 465
Time tag correction 271
Time variable gravity field 288
Topographic reductions 294
Topography 386
Total Least-Squares 417
Troposphere 576
Tropospheric delay 65
Tropospheric path delays 125
Two-step approach 169

Unified height system 32
Upward continuation 214
UT1 119

Variational methods 452
Velocity model 38
Vening Meinesz 392
Vertical component 44
Vertical datum unification 338
Vertical gravity gradients 214
Vertical positioning 21
Very long baseline interferometry 141
Virtual reference stations 71
VLBI 9, 119, 125, 131, 146, 157
Volcano 567
VSOP 137

Wavelet transform 506
Wavelets 320, 392