

The Dynamics of Complex Urban Systems



Sergio Albeverio · Denise Andrey
Paolo Giordano · Alberto Vancheri
(Editors)

The Dynamics of Complex Urban Systems

An Interdisciplinary Approach

Physica-Verlag
A Springer Company

Professor Dr. Dr.h.c. Sergio Albeverio
Institute for Applied Mathematics
University of Bonn
Germany
albeverio@uni-bonn.de
and

Università della Svizzera Italiana
Accademia di Architettura
Via Canavée
6850 Mendrisio
Switzerland
salbeverio@arch.unisi.ch

Denise Andrey
Paolo Giordano
Dr. Alberto Vancheri

Università della Svizzera Italiana
Accademia di Architettura
Via Canavée
6850 Mendrisio
Switzerland

dandrey@arch.unisi.ch
pgiordano@arch.unisi.ch
avancheri@arch.unisi.ch

Library of Congress Control Number: 2007935584

ISBN 978-3-7908-1936-6 Physica-Verlag 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 Physica-Verlag. Violations are liable to prosecution under the German Copyright Law.

Physica-Verlag is a part of Springer Science+Business Media
springer.com

© Physica-Verlag Heidelberg and Accademia di Architettura, Mendrisio, Switzerland, 2008

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.

Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover-design: WMX Design GmbH, Heidelberg

SPIN 11951896 88/3180YL - 5 4 3 2 1 0 Printed on acid-free paper

Preface

In recent years it has become increasingly clear that the dynamics of cities can be best captured by looking at them as complex systems governed by many degrees of freedom, interacting on different space and time scales in a non-linear fashion. The evolution of cities is shaped by internal factors, e.g. decisions taken by institutions and individuals, external causes (international economic context) and by social development. The underlying processes can be slow or fast, acting locally or globally. At a different level, both European cities and megacities are magnets for immigrants (e.g. countries, which have major economical and political problems), leading often to phenomena like marginalization or even ghettization and segregation. They are also hotbeds of economic, political and cultural activity, giving rise, among other things, to relocation and conversion of industries, revaluation of land, and development of new services. These phenomena as *qualitative* changes, are opposed to purely *quantitative* growth processes. They are far from being fully understood, nor they are captured in validated and complete urban models. For most of urban theory hitherto has been based on the assumption of slowly varying spatial and social structures. Only recently, these assumptions have been questioned, giving rise to models employing dissipative dynamics, stochastic cellular automata and multi-agent models, fractal geometry, and evolutionary change models, and to further mathematically oriented approaches. They are promising examples as to how the concepts and methods of mathematics, physics and, more generally natural science can be employed in order to achieve a deeper insight into some aspects of the complexity of urban processes.

One of the aims of the workshop *The Dynamics of Complex Urban Systems: an interdisciplinary Approach* which took place in Ascona (Switzerland), at the “Centro Monte Verità” from 4th to 6th November 2004 was to present and discuss some of these approaches developed by different communities, comparing them and trying to get as much as possible a synthetic overview.

This developed from the conviction that significant progress in the understanding of urban and territorial dynamics can be best achieved through a fruitful collaboration between natural science (physics, mathematics, computer science, biology, ...) and regional science (architecture, geography, city planning, economics, sociology, ...) across traditional disciplines. Different models have to be investigated, which are adapted to various scales and aspects of urban growth. Theoretical but else experimental components are needed for reconstructing and forecasting change processes in cities.

The workshop was structured in 6 sessions characterized by the following key words:

- General dynamical models: urban growth, city evolution, pedestrian dynamics, self-organization, fractal geometry models, urban cluster dynamics, space syntax, continuum state cellular automata

- Models from economics and models for megacities: sustainable urban development, large-scale city formation, sociodynamics, econophysics, synergetics, applied geography
- Models from information science and data management: data mining, soft-computing methods, geo-referenced data, GIS, multi-agents models, artificial worlds, classical cellular automata, data availability
- Related mathematical and physical theories and models: stochastic processes, statistical mechanics, dynamical systems, diffusion, neural networks, power laws, phase transitions
- Models calibration/validation and forecasts: calibration of models' parameters, comparison between empirical data and simulations, optimization, forecasts of stochastic models for complex systems
- Dynamical models and case studies

The interconnections between the sessions turned out in the end to be very strong, which is of course good. This is the reason why for the proceedings we decided to present all contributions in alphabetic order, giving up in particular the division in sections. The contribution by **Michael Batty** was chosen however to open the book since it presents an extensive and deep chronological and conceptual overview of the developments of the urban land use-transportation models which occurred in the last 50 years in the context of current development. The contribution by **Allen, Strathern and Baldwin** presents new models of adaptative organization, which allow a better understanding for integrated views linking land-use changes to environmental and socio-economic and cultural factor. In particular these models stress the importance of adaptable emergent networks. The ontogeny of complex systems models is analyzed in the contribution by **Claes Andersson**, which also discuss the role and applicability of such models. The contribution by **Bazzani, Capriati et al.** investigates the citizen mobility in urban space, presenting an agent-based model for asystemic mobility determined by the "chronotopic areas". The model is also illustrated by a discussion of simulations on the campus of Milano Bicocca University. **Ulrike Beuck, Kai Nagel** et al. present the computational techniques needed for a multi-agent traffic simulation of a metropolitan region as Berlin. Simple behavioural mechanisms in spatial microsimulation models and their dynamic properties are explored by **Mark Birkin**, establishing links between microsimulations (in the contest of a British city), agent-based approaches and spatial interaction models. The difficult of the urban system complexity and the related analysis and forecast discussed by **Ivan Blečić, Armando Cecchini** et al., who in particular try to cope also with the free behaviour of actors. A view of cities as evolutionary systems in random media with spatial emphasis on the intermittency phenomenon is presented in the contribution by **Leonid Bogachev**. The configurational approach to urban phenomena and its further developments are discussed by **Valerio Cutini**, whereas philosophical and methodological issues involved in validation and calibration of cellular automata based models of land use change are discussed in the contribution by **Guy Engelen** and **Roger White**, with illustrations from environmental studies in the Netherlands. Fractal geometry (context, fractal models, morphology and an overview

of results) for modeling of urban patterns is discussed by **Pierre Frankhauser**, with a special attention towards problems of urban sprawl. **Günther Haag** and **Jan Binder** discuss problems of modelling of patterns of a system of different sub-models (e.g. population, transport, production). The theory of the dynamical STASA-model and its application with the SCATTER project for the region of Stuttgart are particularly illustrated. **Erez Hatna** and **Ithzak Benenson** discuss the problem of the appropriateness of modelling urban processes by Markov processes. On the basis of laboratory experiments they argue for shared Markov processes for representing human urban development behaviour and for a basis for decision-making strategies. Three important aspects of self-organization in pedestrian and vehicle traffic are discussed by **Helbing, Johansson** and **Lämmer**. **Jeffrey Johnson** presents a general approach to multidimensional networks as models for complex (urban) systems. The problem of investigating land use transformations, in particular analysing the connections between political and socio-economical changes are discussed by **Silvana Lombardo** and **Massimiliano Petri** on the basis of field investigations in a territorial area of Albania. **Juval Portugali** presents a new structural-cognitive approach to urban simulation models, whereas **Denise Pumain** presents a multi-level model for socio-spatial dynamics of systems of cities and innovation processes. Large scale urban models and their possible renaissance are revisited by **Giovanni Rabino**, who stresses the necessity of joining the scientific and classical cultures in these studies. CityDev a multi-agent simulation of economic spatial dynamics in a poli-nucleated area is discussed by **Ferdinando Semboloni**, and illustrated by a case study in Firenze. The construction and application of continuum-valued cellular automata model for urban systems joined with a fuzzy decision process is presented by **Vancheri, Giordano** et al. In the contribution of **Damiàn Zanette**, Zipf's law relating cities sizes distribution is connected with the theory of multiplicative stochastic processes.

The volume ends with two poster session contributions by Tang Hui Yi and Lu Ming discussing the problems of modelling urban processes in China, in particular in view of their spatial features.

At this point we would like to thank all the speakers for their willingness to contribute to the great success of this conference by their lectures, the interesting discussion during the whole week of the workshop, as well as for delivering the promised manuscripts of their presentations.

We hope that by these proceedings the essential part of the contributions and a little of the stimulating atmosphere of the workshop can be made available for a wide audience.

The editors would like to express their gratitude to the organizers of the workshop professors Michael Batty, Volker Jentsch, Frank Schweitzer and Ferdinando Semboloni for the very stimulating discussions and their help in various stages of the preparation of the workshop and these Proceedings.

In particular the inspiring influence of the International Center of Complex systems (IZKS), University of Bonn, is gratefully acknowledged.

Finally, we would like to thank the *Swiss National Foundation, the Complex Systems Network of Excellence* (Exystence), the *Università della Svizzera Italiana*

(USI) and the *Dipartimento dell'educazione, della cultura e dello sport* (DECS) and its Ufficio degli studi universitari for having supported financially the conference. Many thanks goes also to the staff of the Monte Verità Foundation in Ascona, which with their competence and their kindness have rendered our stay in this marvellous place a real pleasure.

Sergio Albeverio, Denise Andrey, Paolo Giordano and Alberto Vancheri
Spring 2007

Table of Contents

Fifty Years of Urban Modeling: Macro-Statics to Micro-Dynamics <i>M. Batty</i>	1
Complexity: The Integrating Framework for Models of Urban and Regional Systems <i>P.M. Allen, M. Strathern, J. Baldwin</i>	21
Ontogeny and Ontology in Complex Systems Modeling <i>C. Andersson</i>	43
A Model for Asystematic Mobility in Urban Space <i>A. Bazzani, M. Capriotti, B. Giorgini, G. Melchiorre, S. Rambaldi, G. Servizi, G. Turchetti</i>	59
Preliminary Results of a Multi-Agent Traffic Simulation for Berlin <i>U. Beuck, M. Rieser, D. Strippgen, M. Balmer, K. Nagel</i>	75
Hybrid Geographical Models of Urban Spatial Structure and Behaviour <i>M. Birkin</i>	95
Two Complexities and a Few Models <i>I. Bleicic, A. Cecchini, G.A. Trunfio</i>	111
Cities as Evolutionary Systems in Random Media <i>L. Bogachev</i>	143
Grilling the Grid: a Non-Ultimate (Nor Objective) Report on the Configurational Approach to Urban Phenomena <i>V. Cutini</i>	163
Validating and Calibrating Integrated Cellular Automata Based Models of Land Use Change <i>G. Engelen, R. White</i>	185
Fractal Geometry for Measuring and Modeling Urban Patterns <i>P. Frankhauser</i>	213
The Dynamics of Complex Urban Systems: Theory and Application of the STASA-Model within the Scatter Project <i>G. Haag, J. Binder</i>	245
Study of Urban Developers' Behavior in a Game Environment <i>E. Hatna, I. Benenson</i>	265
Self-Organization and Optimization of Pedestrian and Vehicle Traffic in Urban Environments <i>D. Helbing, A. Johansson, S. Lämmer</i>	287

Multidimensional Events in Multilevel Systems <i>J. Johnson</i>	311
The Simulation of Spatial Change: What Relation Between Knowledge and Modeling? A Proposal and Its Application <i>S. Lombardo, M. Petri</i>	335
A Structural-Cognitive Approach to Urban Simulation Models <i>J. Portugali</i>	357
The Socio-Spatial Dynamics of Systems of Cities and Innovation Processes: a Multi-Level Model <i>D. Pumain</i>	373
The Great Return of Large Scale Urban Models: Revival or "Renaissance"? <i>G.A. Rabino</i>	391
The Multi-Agent Simulation of the Economic and Spatial Dynamics of a Poli-Nucleated Urban Area <i>F. Semboloni</i>	409
Modeling the Micro-Dynamics of Urban Systems with Continuum Valued Cellular Automata <i>A. Vancheri, P. Giordano, D. Andrey, S. Albeverio</i>	429
Multiplicative Processes and City Sizes <i>D.H. Zanette</i>	457
<i>Annex: Poster Sessions</i>	
The Features and Challenges of the Urbanization Course in China <i>Tang Hui Yi, Lu Ming</i>	475
Introducing to Simulating Methodology of Urban Systems <i>Tang Hui Yi, Lu Ming</i>	479

Scientific Board

Albeverio Sergio, Prof. Dr. Dr.h.c
Institut für Angewandte Mathematik, Universität Bonn
Wegelerstrasse 6, 53115 Bonn
Germany
albeverio@uni-bonn.de

Batty Michael, Prof. Dr.
Centre for Advanced Spatial Analysis (CASA), University College London
1-19 Torrington Place, London WC1E 6BT
UK
m.batty@ucl.ac.uk

Jentsch Volker, PD Dr.
Interdisziplinäres Zentrum für Komplexe Systeme (IZKS), Universität Bonn
Meckenheimer Allee 176, 53115 Bonn
Germany
jentsch@uni-bonn.de

Schweitzer Frank, Prof. Dr. Dr.
Department of Management, Technology, and Economics (D-MTEC)
Swiss Federal Institute of Technology (ETH)
Zürichbergstrasse 18, 8092 Zürich
Switzerland
fschweitzer@eth.ch

Semboloni Ferdinando, Prof. Dr.
Dipartimento Urbanistica e Pianificazione del Territorio, Università di Firenze
Via Micheli 2, 50121 Firenze
Italy
semboloni@urba.arch.unifi.it

List of Contributors

Albeverio Sergio

Institut für Angewandte Mathematik, Universität Bonn
Wegelerstrasse 6, 53115 Bonn
Germany
albeverio@uni-bonn.de

Allen Peter M.

Complex Systems Management Centre, Cranfield School of Management
Cranfield University
Wharley End, Bedford MK43 OAL
UK
p.m.allen@cranfield.ac.uk

Andersson Claes

Department of Physical Resource Theory, Chalmers University of Technology
41296 Göteborg
Sweden
claes@mindmetric.com

Andrey Denise

Laboratory of Modelling and Applications of Complex Systems (MACS Lab)
Accademia di architettura, Università della svizzera italiana (USI)
Via Canavée 8, 6850 Mendrisio
Switzerland
dandrey@arch.unisi.ch

Baldwin James

Advanced Manufacturing Research Centre, Department of Mechanical
Engineering, University of Sheffield
Sheffield S1 3JD
UK

Balmer Michael

Institute for Transport Planning and Systems (IVT), ETH Zürich
ETH Hönggerberg, 8093 Zürich
Switzerland
balmer@ivt.baug.ethz.ch

Batty Michael

Centre for Advanced Spatial Analysis (CASA), University College London
1-19 Torrington Place, London WC1E 6BT
UK
m.batty@ucl.ac.uk

Bazzani Armando

Laboratory of Fisica della Città, Dept. of Physics, University of Bologna
Via Imerio 46, 40126 Bologna
Italy
armando.bazzani@unibo.it

Benenson Itzhak

ESLab, Porter School of Environmental Studies and Dept. of Geography and
Human Environment, University Tel Aviv
69978 Ramat-Aviv, Tel-Aviv
Israel
bennya@post.tau.ac.il

Beuck Ulrike

Transport Systems Planning and Transport Telematics (VSP), TU Berlin
Salzufer 17-19, 10587 Berlin
Germany
beuck@vsp.tu-berlin.de

Binder Jan

Steinbeis Transfer Centre Applied Systems Analysis (STASA),
Schönbergstrasse 15, 70599 Stuttgart
Germany

Birkin Mark

School of Geography, University of Leeds
Woodhouse Lane, Leeds LS2 9JT
UK
m.birkin@geography.leeds.ac.uk

Blecic Ivan

Laboratory of Analysis and Models for Planning, Dept. of Architecture and
Planning, University of Sassari
Palazzo Pou Salit, Piazza Duomo 6, 07041 Alghero (SS)
Italy
ivan@uniss.it

Bogachev Leonid

Department of Statistics, University of Leeds

Leeds LS2 9JT

UK

bogachev@maths.leeds.ac.uk

Capriotti Massimiliano

Laboratory of Fisica della Città, Dept. of Physics, University of Bologna

Via Irnerio 46, 40126 Bologna

Italy

Cecchini Arnaldo

Laboratory of Analysis and Models for Planning, Dept. of Architecture and Planning, University of Sassari

Palazzo Pou Salit, Piazza Duomo 6, 07041 Alghero (SS)

Italy

cecchini@uniss.it

Cutini Valerio

Dept. of Civil Engineering, University of Pisa

via Diotallevi 2, 56126 Pisa

Italy

valerio.cutini@ing.unipi.it

Engelen Guy

Centre For Integrated Environmental Studies, Flemish Institute for Technological Research (VITO)

Boeretang 200, 2400 Mol

Belgium

guy.engelen@vito.be

Frankhauser Pierre

ThéMA, UMR CNRS 6049, Université de Franche-Comté

32, Rue de Mégevand, 25030 Besançon CEDEX

France

pierre.frankhauser@univ-fcomte.fr

Giordano Paolo

Laboratory of Modelling and Applications of Complex Systems (MACS Lab), Accademia di architettura, Università della svizzera italiana (USI)

Via Canavée 8, 6850 Mendrisio

Switzerland

pgiordano@arch.unisi.ch

Giorgini Bruno

Laboratory of Fisica della Città, Dept. of Physics, University of Bologna
Via Irnerio 46, 40126 Bologna
Italy

Helbing Dirk

Institute for Economics and Traffic, Dresden University of Technology
Andreas-Schubert-Str. 23, 01062 Dresden
Germany
helbing1@vwi.tu-dresden.de

Haag Günter

Steinbeis Transfer Centre Applied Systems Analysis (STASA),
Schönbergstrasse 15, 70599 Stuttgart
Germany
haag@stasa.de

Hatna Erez

ESLab, Porter School of Environmental Studies and Dept. of Geography and
Human Environment, University Tel Aviv
69978 Ramat-Aviv, Tel-Aviv
Israel
erezh51@post.tau.ac.il

Hui Yi Tang

Micro and Macro Architecture, School of Architecture
Harbin Institute of Technology
301 Building C Shengxin tai Sunny Green Garden, Huaihe Rd.
150056 Hangang (HeiLongJiang)
P.R.China
tanghuiyi@126.com

Johansson Anders

Institute for Economics and Traffic, Dresden University of Technology
Andreas-Schubert-Str. 23, 01062 Dresden
Germany

Johnson Jeffrey

Department of Design and Innovation, Faculty of Technology
The Open University
Walton Hall, Milton Keynes MK7 6AA,
UK
j.h.johnson@open.ac.uk

Lämmer Stefan

Institute for Economics and Traffic, Dresden University of Technology
Andreas-Schubert-Str. 23, 01062 Dresden
Germany

Lombardo Silvana

Laboratory of Territorial and Environmental System Engineering (LISTA),
Dept. of Civil Engineering, University of Pisa
via Diotisalvi 2, 56126 Pisa
Italy
s.lombardo@ing.unipi.it

Ming Lu

City Planning and Design Institute, School of Architecture
Harbin Institute of Technology
No. 66 West Da-zhi St., P.O.Box 1534, 150006 Harbin (HeiLongJiang)
P.R.China
hitlm@126.com

Melchiorre Giuseppina

Laboratory of Fisica della Città, Dept. of Physics, University of Bologna
Via Irnerio 46, 40126 Bologna
Italy

Nagel Kai

Transport Systems Planning and Transport Telematics (VSP), TU Berlin
Salzufer 17-19, 10587 Berlin
Germany
nagel@vsp.tu-berlin.de

Petri Massimiliano

Department of Civil Engineering, University of Pisa
via Diotisalvi 2, 56126 Pisa
Italy
m.petri@ing.unipi.it

Portugali Juval

Environmental Simulation Laboratory (ESLab), Dept of Geography and the
Human Environment, Tel Aviv University
69978 Ramat-Aviv, Tel-Aviv
Israel
juval@post.tau.ca.il

Pumain Denise

UMR Géographie-cités, Université Paris I
13 rue du Four, 75 006 Paris
France
pumain@parisgeo.cnrs.fr

Rabino Giovanni A.

Department of Architecture and Planning (DiAP), Polytechnic of Milan
Piazza Leonardo da Vinci 32, 20132 Milan
Italy
giovanni.rabino@polimi.it

Rambaldi Sandro

Laboratory of Fisica della Città, Dept. of Physics, University of Bologna
Via Irnerio 46, 40126 Bologna
Italy

Rieser Marcel

Transport Systems Planning and Transport Telematics (VSP), TU Berlin
Salzufer 17-19, 10587 Berlin
Germany
rieser@vsp.tu-berlin.de

Semboloni Ferdinando

Dipartimento Urbanistica e Pianificazione del Territorio, Università di Firenze
Via Micheli 2, 50121 Firenze
Italy
semboloni@urba.arch.unifi.it

Servizi Graziano

Laboratory of Fisica della Città, Dept. of Physics, University of Bologna
Via Irnerio 46, 40126 Bologna
Italy

Strathern Mark

Complex Systems Management Centre, Cranfield School of Management
Cranfield University
Bedford MK43 OAL
England

Strippgen David

Transport Systems Planning and Transport Telematics (VSP), TU Berlin
Salzufer 17-19, 10587 Berlin
Germany
strippgen@vsp.tu-berlin.de

Trunfio Giuseppe A.

Laboratory of Analysis and Models for Planning, Dept. of Architecture and Planning, University of Sassari
Palazzo Pou Salit, Piazza Duomo 6, 07041 Alghero (SS)
Italy
trunfio@uniss.it

Turchetti Giorgio

Laboratory of Fisica della Città, Dept. of Physics, University of Bologna
Via Irnerio 46, 40126 Bologna
Italy

Vancheri Alberto

Laboratory of Modelling and Applications of Complex Systems (MACS Lab),
Accademia di architettura, Università della svizzera italiana (USI)
Via Canavée 8, 6850 Mendrisio
Switzerland
avancheri@arch.unisi.ch

White Roger

Department of Geography , Memorial University of Newfoundland
St. John's, NF A1B 3X9
Canada
roger@morgan.ucs.mun.ca

Zanette Damiàn Horacio

Consejo Nacional de Investigaciones Científicas y Técnicas
Centro Atomico Bariloche and Instituto Balseiro
Av. Bustillo 9500, 8400 Bariloche, Rio Negro
Argentina
zanette@cab.cnea.gov.ar

Fifty Years of Urban Modeling: Macro-Statics to Micro-Dynamics

Michael Batty

Centre for Advanced Spatial Analysis (CASA), University College London, UK

Abstract

This chapter presents both a chronological and conceptual history of urban land use-transportation models movement in the context of current developments. Such models –‘urban models’ for short – first appeared in the 1950s in North America and were made possible by two interrelated forces: the development of digital computing from which large-scale simulation emanated, and policy imperatives for testing the effects of large-scale public investments on cities. Essentially, urban models are still pragmatically motivated tools for testing the impact of changes in the locations of land use and transportation on dense and usually large urban agglomerations. Planning and policy determine their rationale although their foundations are built on theoretical ideas which go back to the roots of modern social science and the influence of physics and mathematics from the time of the Enlightenment. During the brief but turbulent years since this field has developed, there have been substantial shifts in viewpoint. Indeed even the paradigms that condition what attributes of the city are to be modeled, and the way such modeling takes place, have changed. We will chart these changes, beginning with a set of intersecting time lines focusing on theoretical origins and practical applications. We will show how urban models were first conceived in aggregative, static terms when the concern was for simulating the way cities appeared at a cross-section in time. This aggregative, static conception of urban structure has slowly given way to one where much more detailed disaggregate activities appear more important and where dynamics rather than statics is the focus. This reflects as much our abilities to simulate more elaborate computational structures and collect better data as any grand theoretical revision of the way we look at the city, although such a revision is now under way. As such, this chapter sets a context for many of the current advances in urban modeling reported elsewhere in this book.

1 Historical Antecedents

Wassily Leontieff is best known as the Russian economist who invented the input-output model of the economy in the 1920s before he emigrated to the United States where he subsequently spent his life developing the idea. In contemporary parlance, an input-output model can best be viewed as a large spreadsheet whose

rows and columns represent a standard set of components of the economy such as firms or industries and whose row-column entries give the flow of activity in money or materials from one industry to another. In short, it is a 'flow matrix' that mirrors the interdependencies or linkages between every industry and any other. It is useful largely because if we make the reasonable assumption that the flows between any industry and any other are relatively stable and do not change much in the short term, we can use these dependencies to figure out what would happen in all industries if a single industry or a set of industries grew or declined in size. In fact, although the table contains only the direct effects of such change, it was Leontieff's great contribution to show that one could also figure out the indirect effects, thus linking the idea to the 'multiplier' which featured so strongly in the other macro economic models developed at the same time, particularly those of Keynes.

When Leontieff first formulated his model, he knew full well that he would be able to do very little with it unless he had some rudimentary way of automating the many calculations that such tables required, particularly for figuring out the multiplier effects associated with assessing the impacts of change in and on the economy. In the 1920s and 1930s, when he was developing his ideas at Harvard, Cambridge (Massachusetts) was a ferment of activity involving mechanical computation devices in the years just prior to World War II and the invention of the digital computer. He hooked up with Wilbur Eckert at MIT whose simultaneous equation solver looked up to the task of solving linear systems like input-output models which had a large but nevertheless tractable number of equations and unknowns. To understand the imperatives of those times and why Leontieff found it necessary to engage large scale computation to pursue his quest, we must realize that science and technology had caught the imagination of a very wide public and that progress only seemed assured in the social sciences if we could emulate the unholy liaison that had been fashioned between physics and the development of machine technologies.

In fact with his usual eloquence in recalling those times, Leontieff said that when he sat on the rods that formed the frame of the mechanical equation solver, he could actually change the results of the calculations. The rods were associated with the coefficients of linkage in the input-output table, and sitting on them in different positions amounted to changing their weights which in turn physically changed their value! Tweaking the machine implied tweaking the model and this has become the time honored method of testing the sensitivity of our computer models to change. Our point, of course, is less serendipitous. It is that besides theory, computation was all important to such analysis and certainly essential to its implementation, and that urban models would have never begun and would clearly not be in the form they are today without computation. Analogue soon turned to digital and by the 1950s as soon as computers left the lab and entered commerce in the form of mainframes, engineers and policy makers began to think about ways in which digital conceptions of their systems interest could be used for problem-solving and decision-making. Yet there was still a need for theory. Since the late 19th century, there had been rudimentary but nevertheless insightful attempts at articulating how people located in space in analogy to the way particles and forces

behaved in physics, with action-at-a-distance the all important underlying foundation for why and how we locate in one place or another. Armed with ideas about how gravitation and potential might condition human location, transportation modeling began in the early 1950s closely followed by its extension to embrace land use.

The history of those times is quite well documented (Batty 1979; Harris and Britton 1985; Wegener 1994) but there are three essential issues that have guided the development of urban models ever since. First and foremost, the key driver for this style of modeling has been policy and planning, not a better theoretical understanding of cities. Second, the computational imperative has driven the way these models have been constructed and the way compromises have been made between different model structures in terms of the availability or otherwise of data. Third, what theoretical development there has been has been *ad hoc*. Unlike economic science where there has been a long and deep quest for a theory of the economic system at both micro and macro levels, urban science has developed more pragmatically. Its contributions have come from wide range of different disciplines, many of these being applications of some wider, different theory usually applicable only to partial aspects of the city system. To an extent, this explains why the field is so volatile, dominated by rather different approaches that are hard to reconcile and imply different paradigms and perspectives on what it is that should be explained and modeled.

In the essay that follows, we will begin by outlining various time lines which are composed of several theoretical and practical developments which chart the history of this field over the last 100 years. Our time lines are based on a classification of the various streams which have influenced urban modeling over the last fifty years which is our prime focus but we must supplement it by inquiring what aspects of cities modelers and policy makers have made their major concern. These, as we shall see, have been 'cities-in-equilibrium' whose structure and dynamics were supposed to be explicable in cross-sectional, aggregative ways and we will begin with these. But this paradigm has been found wanting for many reasons. This more than anything else has changed the focus of our field towards the kinds of dynamics that is represented in the majority of chapters in this book. We will then chart this dynamics but parallel our treatment with a foray into questions of detail, of scale, of disaggregation, and the move towards individualistic explanations of urban location and behavior which link our field to complexity theory from the bottom up. We will then illustrate a couple of examples which identify how these various threads are converging and conclude with some speculation that a new form of social physics is in the making, a social physics that now appears much more promising than the classical thinking of fifty years ago but at the same time, a social physics that intrinsically depends on what has gone before.

2 The Time-Lines: Cities, Planning, Modeling

We cannot produce a timeline for a field such as urban modeling without sketching how this fits within what we know and assume about how cities function in terms of urban theory on the one hand, and land use-transportation planning and urban policy making on the other. To this end we will first sketch three related time lines – one for cities, one for planning, and one for modeling before we then elaborate the modeling line which will preoccupy us throughout most of this chapter. Our knowledge of cities is still largely rooted in the way intellectuals and professionals responded to the growth of the industrial city in the 19th century. By and large, cities were seen as being rather stable structures where the dominant functions were located in some central place, or central business district (CBD) as it came to be known in North America. Growth occurred around the periphery and developments in transportation technologies based on energy in the form of the train and automobile reinforced what had been the mono-centric pattern established in ancient and medieval cities around the market place. Some cities did fuse together forming polycentric clusters, conurbations or ‘megalopolis’ as coined by Gottman (1957) but the dominant model was that based on the mono-centre.

Within the city – the so-called intra-urban realm – land uses, social groups and wealth-producing activities appeared to invade and succeed each other according to a simple economic logic. This was generally underpinned by the gradual fall in unit transport costs as technologies improved and as activities found they could consume more space on the edge of the city. In the west at least, the industrial city was one in which the poor were displaced in the centre by employment land use activities while the rich moved to the periphery. This archetypal pattern was consistent with the kinds of segregation and relatively homogenous organization that appeared to exist in many places. Perhaps the clearest statement was from the ‘Urban Ecologists’ writing in the 1920s in Chicago who found this pattern to be the one on which Chicago itself had become organized (Park and Burgess 1925).

Urban planning itself had become institutionalized in the late 19th century to deal with urban problems that were occasioned by the growth of the industrial city. Its instruments were mainly ones of locational control – zoning to avoid the worst excesses of pollution, the preservation of open space through green belts which sought to quite literally ‘stop the city growing’, and decentralization to green field sites in new towns which combined the best of both town and country in terms of quality of life and lower population densities. These strategies contained an implicit reaction against the domination of the single-centered city but to all intents and purposes, they reinforced it by producing new towns and suburbs in the same image, impressing even further the notion that cities should be homogeneously organized in terms of their zoning and land uses. By the 1960s, much of this planning had become explicitly transportation-focused on the provision of infrastructures to keep the pre-industrial urban patterns intact while at the same time engaging in a pattern of urban renewal that reinforced existing zoning. Just as the city was seen as a top-down organization in terms of the way it had developed, so

planning as a function of government was also institutionalized from the top down.

Our ideas about cities and about planning began to shift quite radically during the last quarter of the 20th century. First cities did not appear to be the rather well-organized, homogenous, well-behaved places that had often been assumed earlier. Planning, on the other hand, found itself to be increasingly ineffective in addressing any of the key problem-solving it attempted. This was particularly the case for transportation and for public housing where the problems appeared to get worse rather than better as more intervention in the markets occurred. Planning itself came to be regarded as part of the problem rather than the solution. Cities themselves were increasingly polycentric and although the physical focus was still on the CBD, developments in transportation and in information technology as well as changes in patterns of work through the day and the week loosened the ties to the centre. As ever more populations became urbanized and as the agricultural base shrunk to a tiny fraction of employment, cities began to spread out, merge into one another with cross-commuting becoming the order of the day, replacing the traditional movement from suburbs to down-town. All this was set in an increasingly global world where large cities seemed, indeed are often disconnected from their local hinterlands, and even the nations that contained them no longer seemed relevant to their functioning.

In short, cities appeared to be much more volatile, less stable animals than had hitherto been assumed. The notion that they were homogenous and the fact that they should be planned to be so, was increasingly challenged and the idea that they were dominated by simple patterns of movement and transition through time became passé. Planning itself became more participatory in reaction to the fact that top-down implementation was widely seen as destructive and insensitive. All of this is consistent with what we know about complex systems and although the systems approach had been fashionable in thinking about cities and planning from the 1960s on, a switch from centralized top-down thinking to decentralized bottom-up began to occur on every dimension. The very notion that there was something called a city and something called planning was up for grabs. 'Edge cities' emerged in many parts of the world – cities around cities, and cities within cities – while forecasts that we would all be living in cities by the end of the 21st century, polarized the crisis as to what a city 'actually' was. Our whistle stop tour of the 20th century history of planning and cities is encapsulated within two of the time lines in Fig. 1, but our real focus here is on the kinds of theory and knowledge that was used during these years to fashion the development of urban models that could be used to both explain and predict as well as help inform prescriptions for future cities. It is to these that we now turn.

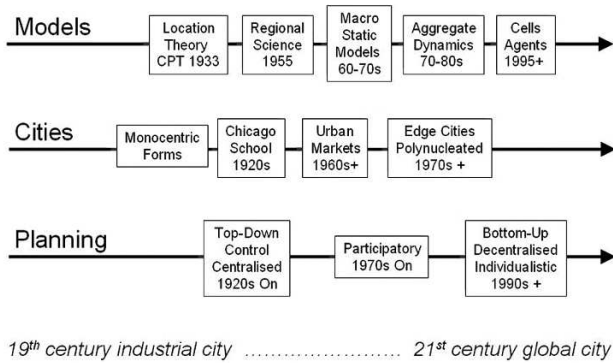


Fig. 1. Intersecting time-line

The theories that were used to underpin our understanding of cities and the various tools that were fashioned to explain and predict their future form closely reflected the top-down, relatively stable, equilibrium-dominated views of planning and cities that we have briefly sketched. Three key ideas of explanation, each based on the notion that it was the cross-sectional structure of cities that should be explained, developed from the late 19th century which we can christen ‘economic location theory’, ‘social physics’, and ‘geographical/spatial morphology’. Location theory emerged somewhat idiosyncratically from the German School in the late 19th century although it was preceded by ideas about rents and markets in the rudimentary economics of Von Thünen and Ricardo in the early part of the same century. This theory essentially argued that industries located according to the balance between their spatial patterns of demand and supply while its generalization to populations sought to show how cities were structured hierarchically from the largest to the smallest according to demand in their hinterlands for the services they provided. This was central place theory developed by Christaller (1933, 1966) in the 1930s and linked to industrial location theory in a coherent economic framework by Losch (1943, 1954) some ten or so years later. It established inter-urban theory based on the idea that systems of cities were also organized spatially as overlapping hierarchical fields while it was picked up by those concerned with the shape or morphology of cities which constitutes our third theme.

Social physics has a longer tradition in that, ever since the late 17th century, there were many ad hoc attempts to apply classical mechanics in the form of Newton’s Laws of Motion to the strength of relationships between people and places at different scales from cities to local neighborhoods (Ball 2004). These were consistent with much of location theory which came later particularly when these theories were treated aggregatively, and it also provided some essential tools to measure proximity, accessibility and to simulate movement between places. The earliest attempt was by Ravenstein who used the gravitational model to explain migration flows in the late 19th century in Britain. It is not the purpose of this chapter to detail the entire history of these movements but readers who wish to get

a sense of this theory should look at Isard's (1956) book *Location and Space Economy* which summarizes all these developments as well as laying out the foundations for regional science that was the bandwagon that pulled all these ideas together in the years following World War 2. Our third theme on geographical/spatial morphology contains both elements of location theory and social physics but the concern is more descriptive, examining ways in which the city is structured. It has been based on the search for patterns in a geographical sense, and generally this corpus of theory has been the domain of urban geographers useful in an operational sense for focusing ideas on what to model rather than how to model the phenomena (Mayer and Kohn 1959).

Once the momentum for fashioning these ideas into tools for urban and transportation planning was established in the 1950s, three distinct sets of techniques emerged to be used as the nuts and bolts from which simulation models were thence developed. Social physics provided the rationale for gravitational models which were used to simulate all kinds of transport flow while micro-economic theories in which the location of individuals and firms could be simulated inside cities as a function of their demand for space, their incomes and their transport costs, were rapidly developed. In this sense, rents and other costs in cities were shown to be inversely related to transport cost or distance, again linking these to the entire gamut of social physics models which were dominated by action-at-a-distance. Models based on the application of macro-economic ideas to the space economy, largely the prerogative of regional science in the form of spatial input-output and econometric forecasting, were also developed, into which more spatially disaggregate models could be embedded. There is no single source covering all these techniques although the book by Isard (1956) and his various successor books cover much of the field while a good summary of ideas from the urban economic standpoint is contained in the book by Fujita (1989) *Urban Economic Theory*.

This then was the context for quite radical changes in urban theory which emerged during the last 25 years of the twentieth century. First, the fact that this entire panoply of models and techniques fashioned around micro-economic theory and social physics treated the city as if it were in equilibrium was questioned from the start. As our collective and documented experiences of how cities change became more complete, it was quite clear that it was growth and change, behavior rather the structure, that was a more appropriate focus for explanation. Second, the notion that surprising things happened in cities had been relegated to an appendix in earlier work but it now appeared that the condition of cities that planning should address is much more bound up with innovation, creativity and surprise than with homogenous land use structures. Third, the idea that cities emerged not from any top-down action but from the bottom up, forced theoreticians and model builders alike to think about emergence, about modeling systems as individuals, not collectives of population and employment. Macro thus moved to micro and static to dynamics. Fourth, the idea of scale came onto the agenda with much planning being concerned more with the small than the large scale. All this was set against massive changes in the computational power and data resources available to those whose concern was urban simulation. Aggregative dynamics dominated develop-

ments in the 1970s and 1980s, and by the early 1990s under the inspiration of complexity theory, urban models based on cells and agents began to appear alongside the long-standing, aggregative, static models of the 1960s. It is now time to unpack these developments in more detail and deconstruct our urban modeling time line shown in Fig. 1 above.

3 Deconstructing the Urban Modeling Time-Line

We could spend this entire book showing how these various time lines can be elaborated, how they emerge, merge, and diverge, how they coalesce and how the key contributors move from one style of theory and model to another. But let us first fine-tune our three perspectives on cities in terms of location theory, social physics, and geographical morphology. As we indicated these three approaches are quite consistent with one another for they reinforce the aggregative, cross-sectional, non-behavioral, non-dynamic view of cities in terms of the theories and their models that we sketched above. But to take our history further we need to show how these foci have provided the momentum for developing much more dynamic, bottom-up disaggregate models of cities which now form the cutting edge of this field and dominate the contributions in this book.

Our urban modeling timeline can be constructed as a composite of these three themes, each representing a line in its own right. Key 19th and first-half 20th century statements of location theory and their subsequent elaboration into regional science and then urban economics are first rooted in the work of Johann Heinrich von Thünen in 1826 in his *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationaloekonomie*. For social physics, we take Walter Christaller's *Die Zentralen Orte in Suddeutschland* in 1933 as the starting point and for spatial morphology we root this in George Kingsley Zipf's *Human Behavior and the Principle of Least Effort* published in 1949. These origins might seem somewhat curious in that Christaller is often associated with location theory and Zipf with social physics. But it is current developments that we have in mind when we make these choices. Location theory remains the purest of these origins for theory rather than modeling, and simulation where social physics remains the focus of this genre. In morphology, theory too remains the focus with an emphasis on aggregative patterns and shape where ideas are used to describe rather than simulate and where the concern is with the physical-spatial properties of cities and regions. It is in the work started by Christaller from which operational models have emerged for it is here that a concern with pattern and analysis, with only looser links to economic processes and aggregative spatial properties, dominate. In short, this second theme is the one from which operational urban models dating back to the 1950s and 1960s. originate although recent developments fuse these traditions in quite subtle ways.

The idea that theory from one or more of these lines is then used in another or the same tradition to design, build and use an operational urban model (which is essentially a computer simulation) is short of the mark. Individuals associated with

these traditions rarely move from one line to another although the influence of the ideas across these themes is strong. Occasionally location theorists have developed optimization models which focus on facility location or urban economic models which link to policy but this is largely because this field is influenced by urban planning and public policy. In contrast, those working in spatial morphology are in more descriptive, less problem-solving oriented traditions and thus most of the models developed in this tradition are largely non-policy-oriented, hence non-operational. Urban models which we will associate with the social physics theme emerged through policy imperatives largely in terms of the coming together of requirements for solving transportation problems in cities in the context of rapidly developing computer technologies which made simulation possible. In the 1950s, social physics ideas were very much in the air and modeling began with the use of the gravity analogue used to simulate flows between origins and destinations where distance or travel cost was the key organizing device reflecting action-at-a distance as implied in central place theory and the rudimentary geography of retailing. By the late 1950s, many of these models were available and hard on their heels came extensions to embrace the location of land use (Voorhees 1959). The watch word of the 1950s in urban planning was that 'transport was a function of land use' and the idea of relaxing trip ends to embrace locational predictions was soon adopted. A flurry of operational models developed in the 1960s culminating in the landmark issue of the 1965 *Journal of the American Institute of Planners* edited by Harris (1965) which represented an excellent early summary of the state-of-the-art.

Most of these early models represented a fusion of social physics ideas with rudimentary regional economics as developed within regional science, the most developed exemplar being Lowry's (1964) *Model of Metropolis*. A variety of simulation techniques were used, ranging from relatively sophisticated econometric analysis to simple event-based simulation. Location theory in so far as it was consistent with gravitational and regional economic modeling, was important but most of the developers of these models, although aware of such theories, did not consider their role as being to implement such theory. Models were built for the purpose at hand against this known but implicit theoretical backcloth. There were some places where there was an appeal to more detailed theory. For example at the University of Pennsylvania in the early 1960s, Alonso's (1964) theory of the housing market which was one of the forerunners of urban economic theory was used to structure a variety of operational models developed by Britton Harris all set against the background of developments in regional science at Penn as well as forming links to new models in the social physics tradition. The same kinds of concern for incorporating economic processes were developed by Kain and his associates at Harvard and although both these developments were inspired by real policy-making, the models were rarely applied in practice (Ingram et al. 1972).

What came out of this experience was a consolidation of techniques with a concern for linking operational models to theory. There was a general feeling amongst theoreticians and model-builders that what was required was much greater consistency concerning model structures and there was a general move to make models ever more comprehensive, embracing more and more urban sectors at ever

more spatial and sectoral detail. For example, in the UK, Wilson (1970) and his colleagues attempted a grand synthesis based on spatial interaction theory which was made consistent using entropy-maximizing analogies with thermodynamic systems in equilibrium. These were also interpreted as part of a wider theory of optimization in which supply and demand within various urban markets could be reconciled with spatial interactions (Wilson et al. 1981). At this time, much stronger economic foundations were laid for such models within discrete choice theory based on utility maximizing (Ben Akiva and Lerman, 1985) while there were attempts to cast these structures within some wider economic equilibrium.

Yet despite the optimism of the 1960s, this was quickly followed by reaction against when models were found wanting along several dimensions. First, in terms of urban planning and policy-making, the models did not address the actual needs for decision support posed by the planners. In short, many models and their model builders sought to answer the wrong questions. When the questions were the right ones, invariably there were arguments over their robustness, given the open and uncertain nature of social prediction while quite often the planning context was so volatile that the very questions changed while the models themselves were still under construction. This was not a good beginning. Combined with the cost of such models and the lack of data along with the fact that this entire domain was being invented on the job, so-to-speak, it is not surprising that the field virtually went into hiding as model-builders retreated to reflect on the experience and nurse their wounds. Lee's (1973) 'Requiem for Large Scale Models' published in the *Journal of the American Institute of Planners* epitomized the vitriol of the reaction.

Those who were reflecting on the models themselves were well aware that cities were much more volatile and heterogeneous affairs than had been assumed hitherto. There was something inconsistent about a domain such as planning which engendered change using models that assumed that the system of interest could move quickly to equilibrium. This was the nature of the theoretical critique but the key problem in articulating models and theories that dealt with urban change rather than urban structure involved our woeful ignorance of urban processes. Moreover the data problem which had plagued the first modeling efforts was doubly severe when it came to thinking about simulating dynamics. Whereas spatial structure was understood to some extent, dynamic structures were much more problematic with little coherent knowledge about how they manifested themselves in cities and certainly little idea about how they impacted on spatial structure. Model builders were forced to look elsewhere for such ideas and as usual it was to physics and mathematics, rather than to the social or biological sciences, that they turned. At much the same time, there were various developments in mathematics focused on rapid and discontinuous change, incorporating radical, qualitative change that became popular. Ideas about how cities could manifest such discontinuous change were examined with catastrophe and bifurcation theory becoming fashionable. When chaos theory became established in the 1980s, these efforts were extended to examine chaotic cycles in urban phenomena. In fact this foray into aggregate dynamics did little for operational modeling which was slowly recovering and adding its own version of urban dynamics by simply re-

peating cross-sectional models at different cross-sections in time. In short by the mid 1980s, the field consisted of the models of the 1960s improved to deal with greater detail with the addition of some quasi dynamics but still being essentially dominated by cross-sectional aggregative statics. In terms of theory, the many forays in aggregate radical dynamics simply served to show how one might proceed but there were few, if any, applications that were developed in practice

From the mid-1980s, however, a sea change began which was not anticipated, indeed had even been regarded as being inconsistent with the way one should theorize about and model any system. It had long been felt that the law of large numbers was an essential underpinning for all science; but what has gradually happened over the last 50 years is a relaxation of these canons of science. When knowledge is always regarded as contingent and never certain as the case with land use-transport models, and when our ability to steer and manage cities is increasingly in doubt, then the structure of a scientific theory that is based on parsimony and generalization comes under severe scrutiny. In short, if the models could not predict anyway, then perhaps the focus should be on building models that informed, extended our understanding, focused us on key issues, but were rich enough to address the questions at hand. Modeling thus began to resemble pedagogy more than prediction, to resemble 'story telling' (Guhathakurta 2002) rather than to provide a definitive understanding of the system of interest and what might happen to it.

In a sense, this sea change in our thinking was paralleled by wider moves to limit the power of government as the grass roots began to reassert itself. Moreover as populations became wealthier, as technologies pervaded all corners of society, then individuals became enfranchised in a way that was very different from the condition of industrial society in the 19th and early 20th centuries. Basically centralization gave way to decentralization with computing technologies and the net being the most potent symbol of this change. In terms of modeling, the focus shifted from aggregates to agents, from groups and collectives to individuals, from large spatial neighborhoods such as census tracts to cells or land parcels, as the quest to model everything in more detail gained the ascendancy. At the same time, the idea of how individuals behaved and their cognition of location and space became more central to the new model styles that emerged. Agent-based modeling and its physical counterpart in cellular automata, as we will discuss in the next section, gradually gained ground but with a very different constituency from that on which the earlier experiences were based. These models are much less rooted in policy and practice and tend to much more speculative than their earlier counterparts. They deal with intrinsic processes of change and in this sense are explicitly disaggregate and dynamic. They embody ideas about how spatial structures might emerge and they have the potential to deal with surprise and innovation. They represent a new way of thinking about cities.

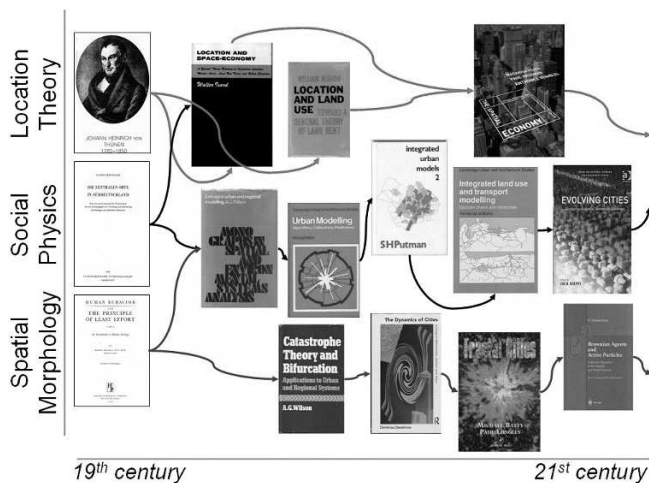


Fig. 2. Snapshots along the urban modeling time line

In parallel to this as our time line implies, the properties of cities have been explored spatially in terms of their morphology through ideas about form and structure using new ideas from geometry (Batty and Longley 1994). The early social physics has also coalesced with these developments and more recently new developments in statistical physics have begun to suggest ways in which cities and related systems can be simulated from the bottom up (Schweitzer 2003). These approaches are beginning to influence operational models as are developments in location theory which incorporate spatial interaction and emergence through growth and trade theory (Fujita et al. 1999). We sketch rather impressionistically these three time lines in Fig. 2 using key statements to identify the stages along the way. These are but snapshots of what has been happening during the period and there are many other similar statements. But this does set the tone for many of contributions in this book where physics-based approaches are rapidly gaining ground as the theoretical rationale for interpreting a variety of structures consistent with these new approaches to simulation.

Before we complete this section, we should not forget that there is also a new wave of land use transportation models – urban models – built around the earlier tradition but being influenced to a degree by these new developments. Many of these models are rooted in behavioral spatial interaction theory and have embraced the agent-based approach but are still largely static in structure, rather than temporally dynamic. Such models are fashioned around developments in transportation modeling and have successfully incorporated new developments in data and GIS in the way they are being constructed (Maguire et al. 2005). There is no real convergence of styles as Fig. 2 implies, nor should there be as different traditions continue to inform one another. But what is clear is that 50 years on from the time when urban models were first applied in the mid 1950s, we now face a much

richer but also much more uncertain style of modeling where the focus is less on predictions, more on understanding and informing. It is to this that we now turn.

4 The Quest for Dynamics: The Macro Perspective

It is hard to know quite why certain concerns and fashions arise in any field and although the quest to make operational models dynamic was widely felt, what began to emerge in the 1970s was quite counter to the most obvious dynamic extensions to existing models. In the 1960s, almost from the inception of land use modeling, some efforts were made to add dynamics by repeating the cross-sectional logic at different points in time but the data problem and the need to produce equilibrium predictions did not elevate this concern to the fore. The first and most dramatic foray into dynamics came from another source. In 1969, Jay Forrester who had applied his ideas about machine dynamics to the firm in the form of industrial dynamics, in talking with the Mayor of Boston, decided to apply his simulation models to the decay of the inner city. What resulted was a model reported only once in his book *Urban Dynamics* (Forrester 1969) but reported in such a way that it made a remarkable splash. Here for the first time was a fully-fledged dynamic model of the city based on the logic of feedback but devoid of any spatial variation whatsoever. This was greeted by the establishment with horror and while it was largely ignored in terms of empirical applications, it did represent a clear statement which others began to emulate in thinking about making the now conventional models dynamic (Batty 1971).

What actually happened in terms of developing ideas about urban dynamics came from elsewhere and in so far as this could be traced to any distinct movement, it came from mathematics, specifically catastrophe theory which had gained the public imagination after Rene Thom published his book *Structural Stability and Morphogenesis* on the subject in 1975. The idea that cities were full of discontinuous change was pursued by Wilson (1981) who largely fashioned his extensions to spatial interaction models in a framework which relied upon non-linear logistic growth, leading to rapid change characteristic of some urban phenomena such as the growth of edge cities and shopping malls. In contrast, Allen (1997) developed Prigogine's ideas associated with reversible thermodynamic systems which illustrated that low-level (local) random change could divert the city's growth path onto very different trajectories, impressing the idea that the actual growth of any city was simply one from an infinite number of possible futures. His model was applied somewhat casually to Brussels while the Wilson models were more focused, particularly on the emergence of retail centers. There were no empirical applications, however, which informed policy.

This concern for dynamics was picked up in the United States as Prigogine and Allen's work was funded by the US Department of Transport and this provided a focus for further work. For example, Dendrinos developed various ideas about dynamics built around coupled non-linear equations of the form used in predator-prey models generating various evolutionary models of the city system at different

scales (Dendrinos and Mullaly 1985). His models although empirically tested, were not applied in a practical problem-solving context with much of the work spinning off from these attempts tending to focus more on theory than practice. Although there was considerable momentum with respect to discontinuity in urban dynamic modeling, this tradition although relating to operational urban models and involving many people with direct links if not experience of building and applying those models, diverged from practice. There were very few attempts at incorporating these ideas in practice. Although the ideas probably made an impact insofar as they alerted the field to the importance of dynamics, the fact that 'cities-in-equilibrium' was not their dominant focus.

By the late 1980s with chaos theory and its relationship to fractal geometry becoming well-known in the sciences, there were various attempts to see such theory as forming a dynamics of city growth and change, following the rather tantalizing possibility that population dynamics could in principle and possibly even in practice lead to chaotic cycling from quite deterministic origins. This gave further weight to Allen and Prigogine's idea that urban growth at the very bottom was essentially dictated by initial conditions that could never be pinned down and that the intrinsic unpredictability of such complex systems was something that should be faced. 'Sensitivity to initial conditions' became the watchword. Dendrinos and Sonis (1990) for example, incorporated much of this theory into their speculations about the dynamic behavior of spatial systems while Nijkamp and Reggiani (1992) provided a useful summary of the state-of-the-art. However, while supporting the general field, this excursion into macro dynamics simply provided a backcloth for discussion and speculation, and did little to extend the art of operational urban modeling useful to policy-makers.

Although the traditional, aggregate, cross-sectional models which formed the origins of the field had been pushed out of practice, this was not for long. Transportation problems in cities and the explosion of urban growth in terms of sprawl was never very far away and some of the key models associated with Putnam (1991) at the University of Pennsylvania, Echenique (1985) at Cambridge and Wegener (2004) at Dortmund continued to be developed. Attention was paid to extending such models to incorporate the local economy in terms of spatial input-output models, interfacing them to more sophisticated discrete-choice spatial interaction transportation models built around the classic four stage process. Disaggregating the model variables also reflected more detail and greater diversity. The notion of repeating the cross-sectional simulations through time was made more transparent, largely because prediction had always been the goal of those modeling efforts. By the 1990s, a reasonable arsenal of practical modeling tools for transport planning and urban growth was available. In fact in the 1980s, a comparative study of several of these modeling efforts in which different models were tested on a standard set of data bases was attempted in the ISGLUTI project (International Study Group on Land Use Transportation Interaction) and although the comparative analysis was limited, this study did detail many of the pitfall and hidden assumptions in constructing and applying such models (Webster et al. 1988).

There were very few new modeling efforts in the original tradition developed during these years but three are worthy of mention: UrbanSim developed by Wad-

dell (2002), the California Urban Futures (CUF) model by Landis and Zhang (1998), and Anas's (1982) residential model based on urban economic markets and transportation choice theory. Waddell's model is the most complete and it alludes to many new ideas in modeling urban systems, has well-developed urban economic and transportation components and is quasi-dynamic. Anas's model is not dissimilar although it is restricted to the residential and transportation sectors, while Landis and Zhang's model is focused on the land development process. Recently there have been quite impressive attempts to link several of the key operational models to various kinds of environments and evaluation indicators. In the EU SPARTACUS project and its follow-up PROPOLIS, traditional model structures are being extended to embrace GIS through common interfaces for the MEPLAN, TRANUS (de la Barra 1989) and Dortmund Models (Spartacus Consortium 1998; Propolis Consortium 2004; Wegener 2004). Currently there is added interest in traditional models as many of them have painfully struggled to embrace new technologies and ideologies while keeping their operational focus.

5 Towards Micro Dynamics: Agents, Cells and the New Social Physics

Somewhere along the way, new currents have emerged from both within the field and without. From within, the focus began to change due to the existence of new data sets and new technologies such as GIS and a new generation of physical models have appeared. These attempt to simulate the development process in terms of urban growth, building on simple ideas about diffusion as in cellular automata (CA) modeling. This has been paralleled by a concern for even finer-scale disaggregation for simulating populations at their most individualistic level. This is not micro-simulation in the traditional sense as it has been used in social and urban simulation (Clarke 1996) but the representation and simulation of individual agents in terms of their preferences and movement patterns. These agent-based models are much wider than urban simulation *per se*. A lot of social science and some physical science have been pervaded by the 'agent-based' viewpoint largely because computer systems and fine-scale data have made such representations possible. Agent-based models, in a sense, are being informed from without although geodemographics – the study of fine scale population profiles at a highly disaggregate group and spatial scales – is forcing their development.

The wider context for both cellular automata modeling of urban development processes and agent-based models of urban movement and change clearly reflects a new view of systems theory – complexity theory – which has shifted the long-lasting concern for the city as a system from its structure to its behavior. Dynamics has come back firmly onto the agenda embracing not only the macro-dynamics of discontinuity discussed above but also micro-dynamics posed by individuals operating from the bottom up. This shift from the top-down has occasioned much speculation about cities being emergent systems where the structures evolved are often novel and surprising, thus relating to Allen and Prigogine's ideas about urban

futures which were first spelt out two or more generations ago. At the same time, there has been a flowering of new ideas and techniques in statistical physics. In particular the notions of critical thresholds and of systems far-from-equilibrium have gained much ground while new techniques to study diffusion and the growth of network structures has fashioned a new generation of models which might be said to represent a new social physics (Schweitzer 2003, Batty 2005a; Andersson 2005). From a rather different angle has come a new view of the spatial economy which is closely linked to many of these ideas in the new physics where dynamics, movement and trade are central (Fujita et al. 1999). In essence, what these new theories are providing is a coherent set of ideas about dynamics at the individual as well as collective level with clear links to the spatial morphologies generated, and linked in various ways to long standing traditions in location theory and urban economics. Everywhere one looks, there are concerns for merging these various traditions as Fig. 2 implies. Indeed at this point in time, there is a great melting pot of ideas being used to construct different perspectives on the urban system.

Very few of these models are operational in the traditional sense. But in a way, this is consistent with the retreat from prediction that has become more acceptable even in a practical context where the need for discussion of the urban future is still as urgent as ever. Modeling as story telling, as pedagogy, as informed speculation, has become the rationale for the development of these new ideas and it is still too early yet to see how it will all pan out. The chapters in this book are as significant as barometer as we currently have. To conclude, let us focus a little on those models which are closer to operationality and we will take cellular automata as our exemplar. Cell-based models in analogy with CA were suggested for urban growth as far back as the early 1960s and in that tradition, the models developed by the Chapin and Weiss (1968) in North Carolina and by Lathrop and Hamburg (1965) in Buffalo, NY are noteworthy. In 1970, Tobler developed his own simulation of the growth of Detroit and through the 1970s, he speculated on how cell-based ideas might be used to simulate diffusion in local spatial neighborhoods (Tobler 1970, 1979). Couclelis (1985) continued this tradition in the 1980s and the growth of GIS gave the area a push with raster-based data sets often being used for cellular representation. What is noteworthy about this emerging tradition is that it is not really strongly linked to operational land use-transportation modeling except through individuals who have worked on both or been associated with both. It is mainly geographers who have developed CA models and their focus has been almost entirely physical. The major critique of this work relates to the absence of an urban economy underpinning most of the models, and the somewhat cavalier approach to action-at-distance which is subsumed within local neighborhood effects and transitions. By and large, transportation is missing from these models. Where these limitations have been relaxed, the models appear quite close to some of those developed 40 years or more ago.

Very few of these CA models have been developed for practical policy applications. Batty (2005b) provides a list of the main groups working in this area. The two that have got closest to practical applications are the Santa Barbara group (Clarke et al. 1997) and the group led by White and Engelen (1997) at RIKS. Both groups have developed hands-off type policy applications mainly for regional and

national agencies: the Santa Barbara group for urban growth in the US for USGS (United States Geological Service) while the RIKS group for the European Commission. Again where the model structures are relaxed, they come closer to the traditional corpus of land use-transportation models. There are now probably upwards of 50 such applications world-wide but most exist in academia and relate more strongly to the new traditions in complexity theory and social physics than they do to the ongoing urban modeling tradition. In terms of the development of agent-based equivalents, although there are many partial models of urban sectors dealing with processes such as residential segregation, pedestrian movement and so on, there are hardly any constructed for urban development. These are of course in their infancy and it is entirely possible that in the next decade there will be attempts linking the largely non-spatial micro-simulation of cities to their spatial equivalents which will be agent-based. Several of the chapters that follow imply these developments.

6 What Has Been Achieved: Retrospect and Prospect

One of fascinating features of this 50 year history is that most of the people who contributed to it are still alive and if not working within it, are conscious of how it has and is developing. Although we have sought to show how different lines of development have been generated spontaneously or have emerged naturally from developments so far, bringing in new people or developing the expertise of those already in the field, the field is still quite narrow and relatively focused. As one moves more towards practical applications, there is less concern for new theoretical developments but as urban models are still so idiosyncratic in their design and construction, most model builders have to be and are indeed aware of the general state-of-the-art. In this sense, the field is still coherent and tight; it is the context that has changed. The models built 50 years ago were constructed on-the-fly, so-to-speak, in practice, and when viewed from the vantage point of the early 21st century, it was something close to a miracle that they worked at all. In perspective, although the experience was salutary and the first generation of models and modelers were exposed to a baptism of fire, the field continued largely because the policy problems that motivate the need for a better understanding of cities and the need to predict the future have not gone away.

What has changed is our perspective on what is possible. Simulation itself is no longer just about predicting the right future but about predicting many futures. Modeling is about story telling, about informing us of many possible futures. And all this is consistent with the notion that cities and the societies they are a part of are intrinsically complex and inherently unpredictable. The question as to why we still need predictive models in this context is still not resolved for there are many who still consider that this way of thinking is a luxury society cannot afford. But slowly and surely, the view is gaining ground that the informed speculation that such simulation clearly brings and the ability to communicate this through models is a valuable focusing activity; and if only for this, many practical agencies man-

dated to grapple with urban problems accept the need for this style of modeling. In short, what has emerged is a hierarchy of model types, and it is fitness for purpose that is now the distinguishing mark that must be applied when considering applications. These themes are picked up time and again in the chapters in this book. The conundrums and paradoxes of a complex urban world will always remain but our ability to handle them is surely informed and extended by the new generation models presented here.

References

- Allen PM (1997) *Cities and Regions as Self-Organizing Systems: Models of Complexity*. Taylor and Francis, London
- Alonso W (1964) *Location and Land Use*. Harvard University Press, Cambridge, MA
- Anas A (1982) *Residential Location Markets and Urban Transportation: Economic Theory, Econometrics and Policy Analysis with Discrete Choice Models*. Academic Press, New York
- Andersson C (2005) *Urban Evolution*. Doctoral thesis, Chalmers Institute of Technology, Göteborg, Sweden
- Ball P (2004) *Critical Mass: How One Thing Leads to Another*. Heinemann, London
- Batty M (1971) Modelling Cities as Dynamic Systems. *Nature* **231**: 425-428
- Batty M (1979) Progress, Success and Failure in Urban Modelling. *Environment and Planning A* **11**: 863-878
- Batty M (2005a) *Cities and Complexity*. MIT Press, Cambridge, MA
- Batty M (2005b) Agents, Cells, and Cities: New Representational Models for Simulating Multi-scale Urban Dynamics. *Environment and Planning A* **37**, in press
- Batty M, Longley PA (1994) *Fractal Cities: A Geometry of Form and Function*. Academic Press, San Diego, CA
- Ben Akiva M, Lerman S (1985) *Discrete Choice Analysis*. MIT Press, Cambridge, MA
- Chapin FS, Weiss SF (1968) A Probabilistic Model for Residential Growth. *Transportation Research* **2**: 375-390
- Christaller W (1966) *Central Places in Southern Germany*. Prentice-Hall, Englewood Cliffs, NJ (translated from the 1933 German Edition *Die Zentralen Orte in Sueddeutschland* by CW Baskin)
- Clarke G (ed) (1996) *Microsimulation for Urban and Regional Policy Analysis*. Pion Press, London
- Clarke KC, Hoppen S, Gaydos L (1997) A Self-Modifying Cellular Automaton Model of Historical Urbanization in the San Francisco Bay Area. *Environment and Planning B* **24**: 247-261
- Couclelis H (1985) Cellular Worlds: A Framework for Modeling Micro-Macro Dynamics. *Environment and Planning A* **17**: 585-596
- de la Barra T (1989) *Integrated Land Use and Transport Modelling*. Cambridge University Press, Cambridge
- Dendrinis DS, Mullally H (1985) *Urban Evolution: Studies in the Mathematical Ecology of Cities*. Oxford University Press, Oxford, UK
- Dendrinis DS, Sonis M (1990) *Chaos and Socio-Spatial Dynamics*. Springer, New York

- Echenique M (1985) The Use of Integrated Land Use and Transport Models: The Cases of Sao Paulo, Brazil and Bilbao, Spain. In: Florian M (ed) *The Practice of Transportation Planning*. Elsevier, Amsterdam
- Forrester JW (1969) *Urban Dynamics*. MIT Press, Cambridge, MA
- Fujita M (1989) *Urban Economic Theory: Land Use and City Size*. Cambridge University Press, Cambridge, UK
- Fujita M, Krugman P, Venables AJ (1999) *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Cambridge, MA
- Gottman J (1957) Megalopolis, or the Urbanisation of the North Eastern Seaboard. *Economic Geography* **33**: 189-200
- Guhathakurta S (2002) Urban Modeling as Storytelling: Using Simulation Models as Narrative, *Environment and Planning B* **29**: 895-911
- Harris B (1965) Urban Development Models: A New Tool for Planners. *Journal of the American Institute of Planners* **31**: 90-95
- Harris B, Britton B (1985) Urban Simulation Models in Regional Science. *Journal of Regional Science* **25**: 548-568
- Ingram GK, Kain JF, Ginn JR (1972) *The Detroit Prototype of the NBER Urban Simulation Model*. National Bureau of Economic Research, Columbia University Press, New York
- Isard W (1956) *Location and Space-Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, Trade and Urban Structure*. MIT Press, Cambridge, MA
- Landis JD, Zhang M (1998) The Second Generation of the California Urban Futures Model. Part 1: Model Logic and Theory. *Environment and Planning B* **25**: 657-666
- Lathrop GT, Hamburg JR (1965) An Opportunity-Accessibility Model for Allocating Regional Growth. *Journal of the American Institute of Planners* **31**: 95-103
- Lee DB (1973) Requiem for Large Scale Models. *Journal of the American Institute of Planners* **39**: 163-178
- Losch A (1954) *The Economics of Location*. Yale University Press, New Haven, CN (translated from the 1943 German Edition *Die Raumlische Ordnung der Wirtschaft* by WH Woglom)
- Lowry IS (1964) *Model of Metropolis*. Memorandum RM-4035-RC, Rand Corporation, Santa Monica, CA
- Maguire D, Batty M, Goodchild M (eds)(2005) *GIS, Spatial Analysis, and Modeling*. ESRI Press, Redlands, CA
- Mayer HM, Kohn CF (eds) (1959) *Readings in Urban Geography*. University of Chicago Press, Chicago, Illinois
- Nijkamp P, Reggiani A (1992) *Interaction, Evolution and Chaos in Space*. Springer, Berlin
- Park RE, Burgess EW (1925) *The City*. University of Chicago Press, Chicago, Illinois
- Portugali J (2000) *Self-Organization and the City*. Springer, Berlin
- Propolis Consortium (2004) *PROPOLIS (Policies and Research of Policies for Land Use and Transport for Increasing Urban Sustainability)*. Final Report for the Commission of the European Communities, from LT Consultants Ltd., Helsinki, Finland
- Putman SH (1991) *Integrated Urban Models 2. New Research and Applications of Optimization and Dynamics*. Pion, London, UK
- Schweitzer F (2003) *Brownian Agents and Active Particles: Collective Dynamics in the Natural and Social Sciences*. Springer, Berlin

- Spartacus Consortium (1998) SPARTACUS (System for Planning and Research in Towns and Cities for Urban Sustainability). Final Report for the Commission of the European Communities, from LT Consultants Ltd., Helsinki, Finland
- Thom R (1975) Structural Stability and Morphogenesis. Benjamin, Reading, MA
- Tobler WR (1970) A Computer Movie Simulating Population Growth in the Detroit Region. *Economic Geography* **42**: 234-240
- Tobler WR (1979) Cellular Geography. In Gale S, Olsson G (eds) *Philosophy in Geography*. Reidel, Dordrecht, pp 279-386
- von Thünen JH (1966) Von Thünen's Isolated State. Pergamon, Oxford, UK (translation from the 1826 German Edition *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationaloekonomie* by PG Hall)
- Voorhees AM (1959) Land Use and Traffic Models. *Journal of the American Institute of Planners* **25**: 55-57
- Waddell P (2002) UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning. *Journal of the American Planning Association* **68**: 297-314
- Webster FV, Bly P, Paulley NJ (eds)(1988) *Urban Land Use and Transport: Policies and Models*. Avebury, Aldershot, UK
- Wegener M (1994) Operational Urban Models: State of the Art. *Journal of the American Planning Association* **60**: 17-29
- Wegener M (2004) Overview of land-use transport models. In: Henschler DA, Button K (eds) *Transport Geography and Spatial Systems. Handbook 5 of the Handbook in Transport*, Pergamon/Elsevier Science, Kidlington, UK
- White RW, Engelen G (1997) Cellular Automaton as the Basis of Integrated Dynamic Regional Modelling. *Environment and Planning B* **24**: 235-246
- Wilson AG (1970) *Entropy in Urban and Regional Modelling*. Pion Press, London
- Wilson AG (1981) *Catastrophe Theory and Bifurcation: Applications to Urban and Regional Systems*. University of California Press, Berkeley, CA
- Wilson AG, Coelho JD, Macgill SM, Williams HCWL (1981) *Optimization in Locational and Transport Analysis*. John Wiley, Chichester, UK
- Zipf GK (1949) *Human Behavior and The Principle of Least Effort*. Addison-Wesley, Cambridge, MA

Complexity: the Integrating Framework for Models of Urban and Regional Systems

Peter M.Allen¹, Mark Strathern¹, James Baldwin²

¹ Complex Systems Management Centre, Cranfield University, UK

² Advanced Manufacturing Research Centre, University of Sheffield, UK

Abstract

Traditionally, science has attempted to understand urban systems using a reductionist approach in which the behaviour of a system (city or region) is represented as being an equilibrium, mechanical interaction of its components. These components are “representative agents” for the different categories of supply and demand that inhabit the system, and it is assumed that their spatial distribution reflects an optimised value of profit (supply) and utility (demand). Over recent decades many attempts have been made to introduce more dynamic approaches, in which equilibrium is not assumed, and there are many models and methods that attempt to do this. However, this still denies the essential complexity of the urban or regional system, in which activities, natural endowments, culture, skills, education, health, transport, house prices, the global economy, all combine to affect the evolution of the system. Just as in ecology, the key to the long-term structures that may emerge is the diversity, innovative and adaptive power of people and society to counter new difficulties and create new opportunities. This fluid, adaptive power is a product of the complex system, and can only be modelled and anticipated to a limited degree. However, cities and regions can limit the possibility of successful adaptation if they are too “well-organized” or too unimaginative. New models of adaptive organisation allow us to understand better the need for integrated views linking land-use changes to environmental and socio-economic and cultural factors. These provide a new, more open way of considering the importance of adaptable, emergent networks, and the need for multiple and burgeoning accessibility to others.

1 Introduction

In order to make effective designs, investments and policy decisions in cities, we need to understand the multiple decisions and actions made by the multiplicity of agents and entities involved. This really means that we need to understand the options that they perceive, and the trade-offs that their value systems cause them to make, and through this to know how they will react to some policy, action or investment that is contemplated. Only then would we have a reasonable basis on

which to identify emerging problems and to evaluate different possible policy or decision responses.

In fact, the behaviour of *complex systems* offers an appropriate set of concepts with which to begin a new reflection on human systems. In this new view, *non-equilibrium* phenomena are much more important, and offer a new understanding of the natural emergence of structure and organization in systems with many interacting individual elements. In this paper we briefly refer to new models of evolutionary and regional systems that show how the dialogue between the individual and collective levels generate successive spatial structures, with characteristic patterns and flows. These represent a co-evolutionary behaviour and organization beyond the "mechanical" where, the locations and behaviours of the actors are mutually inter-dependent, the system has many possible responses to perturbations, and where the urban system can change, adapt and maintain rich, diverse and varied strategies. This view of sub-optimal behaviours, imperfect information, mistaken inferences and the power of creativity is contrasted with the traditional mechanical representations of human systems. The models discussed here offer a new, quantitative basis for policy exploration and analysis, allowing us to take into account the longer-term implications for the system as a whole.

2 Evolutionary Complex Systems

In a series of articles the fundamental properties of complex systems have been established (Allen 1983, 1990, 1997a). Let us consider the most basic problem of modelling a natural ecosystem. We can establish the different species that exist there, and then find out how many of each population there are. We can also, by sampling, find out which population eats which other population and calibrate the multiple plant/herbivore and predator/prey interactions. Now, once this is established, we can put the whole system of equations on a computer, and run it forward. What happens is shown in Fig. 1.

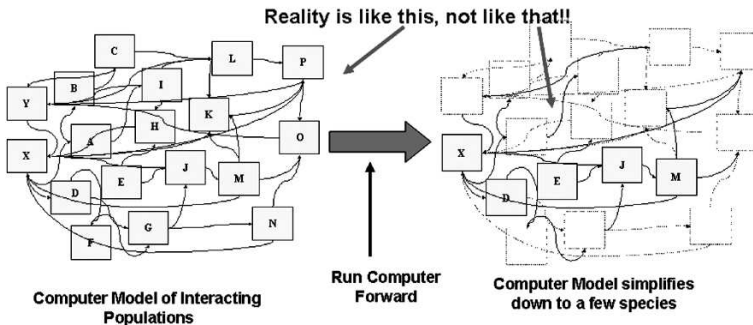


Fig. 1. A calibrated ecosystem represented by the population dynamics of its constituent species collapses when run forward in time

The model collapses down to a simple, very reduced structure. This is an astonishing result. It means that although the model was calibrated on what was happening at time $t = 0$ it diverged from reality as time moved forward. The real ecosystem stayed complex, and indeed continued to adapt and change with its real environment. But this shows us that the mechanical representation of reality differs critically from that reality. What is missing? This can be discovered if we examine carefully the assumptions that we made in formulating our population dynamics. What happened is that the loops interactions of a real ecosystem form parallel food chains, with cross connections and complications of course, but essentially with each level feeding on the lower one, some of these dying and others being eaten by the level above. The whole system of food chains loops back through death and micro-organisms that recycle all the carbon and minerals. When we run the population dynamics with the fixed birth, death capture and escape rates that we have found on average in the real system (in analogy with chemical reaction rates), then the food chain with the highest performance simply eliminates all the others. In other words, selection between metabolic chains operates and this selects for the highest performing chain. However, reality does not. Therefore we need to understand what is missing between the dynamic model and the original real system.

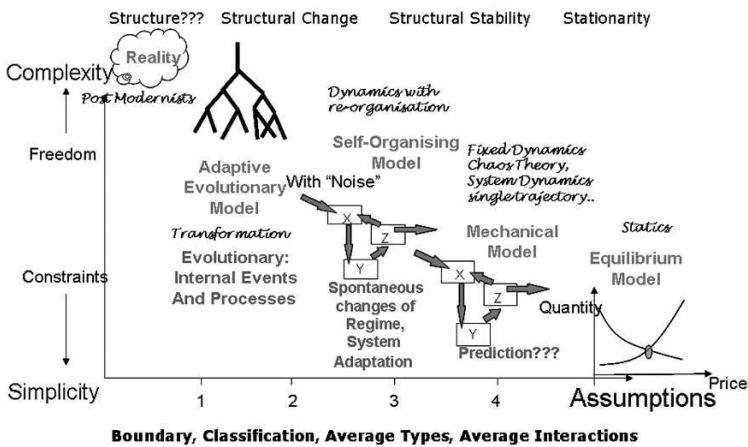


Fig. 2. This shows the results of successive simplifying assumptions that take us from a complex evolving system to its mechanical representation.

The key answer is that what is missing is the *internal diversity of the populations*. In chemistry, one molecule is very like another, and the only difference is their spatial location. Dissipative structures can create spatio-temporal patterns because of this. But populations of organisms differ in an infinite number of ways. Firstly in location, but also in age, size, strength, speed, colour etc. and so this means that whenever a population, X, is being decreased by the action of some particular predator or environmental change, then the individuals that are most

vulnerable will be the ones that “go” first. Because of this the parameter representing the average death rate will actually change its value as the distribution within the population X increases the average “resistance”. In other words, the whole system of populations has built in through the internal diversities of its populations, a multiple set of self-regulatory processes that will automatically strengthen the weak, and weaken the strong. In the same way that reaction diffusion systems in chemistry can create patterns in space and time, so in this more complex system, the dynamics will create patterns in the different dimensions of diversity that the populations inhabit. But neither we, nor the populations concerned, need to know what these dimensions are. It just happens as a result of evolutionary dynamics.

In this case it becomes key to understand the sequence of assumptions that take us from reality to a mechanical representation of that reality. This leads us to the general view that is shown in Fig. 2. This sets out the kind of models that result from a particular set of assumptions.

Number	Assumption Made	Resulting Model
1	Boundary assumed	Some local sense-making possible – no structure supposed
2	Classification assumed	Open-ended Evolutionary models – structural change occurs
3	Average Types	Probabilistic, Non-Linear Equations – assumed structurally stable
4	Average events	Deterministic, Mechanical Equations – assumed structurally

This succession of models arises from making successive, simplifying assumptions, and therefore models on the right are increasingly easy to understand and picture, but increasingly far from reality. *They also are shorn of their capacity to evolve – their real underlying exploratory, error-making processes.* The operation of a mechanical system may be easy to understand but that simplicity has assumed away the more complex sources of its ability to adapt and change. They become more like “descriptions” of the system at a particular moment, but do not contain the magic ingredient of micro-diversity that will really allow the system to undergo structural change and create a new, qualitatively different system, with some new variables and some emergent performance. The ability to adapt and change is still present in the “evolutionary” model that only makes assumptions 1 and 2, but not those of average type and average behaviours. This therefore tells us that the evolutionary capacity is generated by the behaviours that are averaged by assumptions 3 and 4 – average types and average events – and therefore that organisations or individuals that can adapt and transform themselves, do so as a result of the generation of micro-diversity and the interactions with micro-contextualities. This tells us the difference between a reality that is “becoming” and our simplified understanding of this that is merely “being” (Prigogine 1980).

In reality, complex systems thinking offers us a new, integrative paradigm, in which we retain the fact of multiple subjectivities, and of differing perceptions and views, and indeed see this as part of the complexity, and a source of creative inter-

action and of innovation and change. The underlying paradox is that knowledge of any particular discipline will necessarily imply “a lack of knowledge” of other aspects. But all the different disciplines and domains of “knowledge” will interact through reality – and so actions based on any particular domain of knowledge, although seemingly rational and consistent, will necessarily be inadequate.

Management, or policy exploration require an integrated view. These new ideas encompass evolutionary processes in general, and apply to the social, cultural, economic, technological, psychological and philosophical aspects of our realities. Often, we restrict our studies to only the “economic” aspects of a situation, with accompanying numbers, but we should not forget that we may be looking at very “lagged” indicators of other phenomena involving people, emotions, relationships, and intuitions – to mention but a few. We may need to be careful in thinking that our views will be useful if they are based on observations and theories that refer only to a small sub-space of reality – the economic zone.

The underlying causes and explanations may involve other factors entirely, and the economic “effects” of these may be only delayed, ripples or possibly tidal waves. What matters over time is the expansion of any system into new dimensions and conceptual spaces, as a result of successive instabilities involving dimensions additional to those the current “system” appears to occupy.

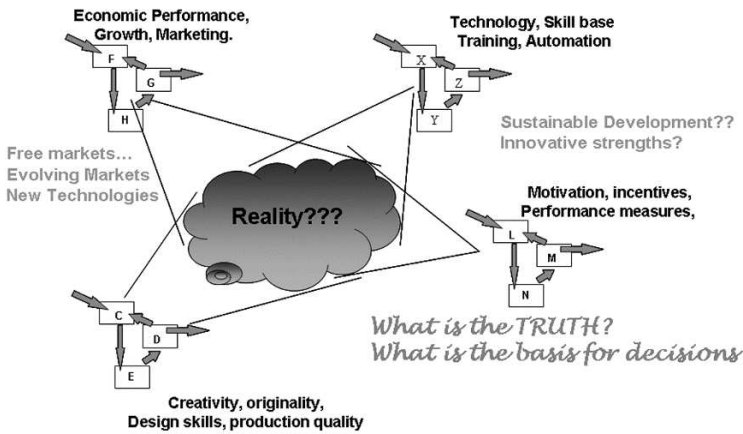


Fig. 3. Different people see the same system in different ways. Each can however be rational and consistent, whilst advocating quite different actions or policies.

This idea of evolution as a question of “invasibility”, with respect to what was not yet in the system, was the subject of a very early paper by the author (Allen 1976). Essentially then, systems are seen as temporary, emergent structures that result from the self-reinforcing non-linear interactions that result from successive “invasions”. History is written not only by some process of “rational improvement” in its internal structure but more fundamentally by its dialogue with ele-

ments that are *not yet* in the system – successive experimental linkages that either are rejected by the system, or which “take off” and modify the system irreversibly.

Rational improvement of internal structure, the traditional domain of “systems’ thinking”, supposes that the system has a purpose, and known measures of “performance” which can indicate the direction of improvements. But, this more fundamental structural evolution of complex systems that results from successive invasions of the system by new elements and entities, is characterized by emergent properties and effects, that lead to new attributes, purposes and performance measures. In the next sections therefore, we attempt to show that this structural evolution is not in fact “random” in its outcome, as successful invasions of a system are always characterized by the revelation of positive feedback and synergy, creating particular new, internally coherent, structures from a growing, explosively rich set of diverse possibilities.

The realm of “complex systems” models that we wish to develop aim to make only the first two assumptions, and to study cities and regions as evolving, self-transforming systems in which behaviour, decisions and the value systems underlying these all evolve over time. This leads to a view of a city or region as a complex evolution of spatially distributed learning reflecting local stresses, opportunities and exploratory responses such that people not only change what they do, but also their knowledge of what they could do, and what they want to do. Qualitative, structural changes occur both in the macroscopic forms of the collective structure, and also in the microscopic structures within individuals’ brains that govern their trade-offs and decision making, which in turn govern the future structural evolution of the collective system and of the individuals that inhabit it.

Another key input to the paper here is based on recent research concerning the evolution of organisational forms in business. In particular, the historical evolution of manufacturing companies has been analysed in terms of their “constituent working practices” and an “evolutionary tree” of organisational forms deduced (McKelvey, Ridgway and McCarthy). This corresponds to a kind of historical picture of successive steps in the division of labour as new techniques, practices and types of skill emerged. Further to this, in Allen et al. (2005) it has been shown that the organisational forms that have successfully emerged over time, correspond to particular structures in which the constituent practices, skills and techniques all fit together in a synergetic pattern of positive feedback. Such temporary islands of functional a survey of manufacturers allowed.

In reality then a city is a complex system, as is a neighbourhood, a block, a household, and an individual. These represent nested levels of description, and we can develop mathematical models that can explore different possible evolutionary pathways and possible futures under the assumptions of different possible interventions. This work started in 1976 when the US Department of Transportation commissioned our early research on developing such models, initially only dynamic, but later on developing fully complex, “learning” multi-agent models. The essence of these models is shown in Fig. 4 in which the locational patterns of people, of jobs, of transport and of infrastructure are coupled together, so that their combined evolution can be explored under different interventions and plans. The key idea is that changing transport costs or access change the locational choices of

people and activities, that in turn change the patterns of traffic flow and the locational choices of people and activities, and so on. These strong feedback relations mean that the system is unstable, and can exhibit different possible trajectories into the future. In examining possible decisions therefore, we need to see whether the different possible trajectories of the system are all acceptable, or whether there is a probability that some very unpleasant outcomes might be possible.

An important point to underline concerns the reasons for which we travel at all.

If we ask why people travel at all, we find that it is because of the spatial distribution of diverse activities and opportunities:

- Dispersed distribution of affordable/desirable housing
- Concentrated distributions of employment
- Concentrated distributions of retail opportunities
- Dispersed distributions of leisure and cultural facilities

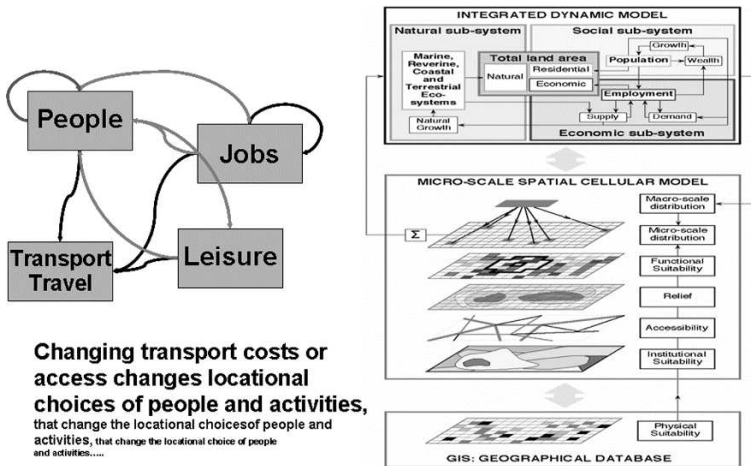


Fig. 4. Software systems have been developed that allow the interacting spatial distributions of people, jobs, leisure facilities and transportation can be studied. (White and Engelen 2001)

Transport demand is therefore generated by these spatial distributions, and an important point is that these distributions are all CO-EVOLVING with each other over time, and also reflect changes in the transportation systems. In short then, the demand for transport is generated by the details of the different distributions, that affect each other, and the transport congestions and patterns of access then affect the locations of the different spatial distributions, which in turn feed back on the demand for transport. This is particularly evident at the present time in the UK when spatially dependent house price rises are currently shaping longer commuting patterns for large numbers of people, and even threatening the successful functioning of cities, as ordinary workers, particularly in the public sector, find it increasingly difficult to find homes within reasonable distances.

In view of this complexity, and the intertwined effects of transportation and spatial structure, it seems clear that there is a problem for the evaluation of transportation policies and plans. How can there be an overall assessment or evaluation of any plan for new roads or for public transport systems unless these complex effects are assessed? The answer is that, in fact, they are not. Decisions concerning urban highways, new tram and metro systems are really decided on the basis of politics and fashion.

3 Dynamic, Spatial Urban Models

Since the 1970s, work has been going on that attempted to develop computer models that would take into account the complex interactions of linked responses that lead to a co-evolution of urban structure (patterns of retail, commercial and manufacturing employment, and different qualities of residence) with transportation infrastructure. These models are based on the following characteristics:

- Different types of actor at each zone, with characteristic needs to be fulfilled
- These characteristic needs are stable, but the behaviour of actors depends on the changing circumstances
- The spatial distributions of the different types of job and different kinds of people affect each other as the potential for housing demands, commercial activities and for travel affect and are affected by transportation and land-use.

The development of these models has been described in “Cities and Regions as Self-Organizing Systems” (Allen 1997b). After an initial phase that developed models suitable for some US and European cities, an example based on Brussels was developed to demonstrate the potential utility of the approach. The model represents the interacting behaviours of the actors in the urban system, as they each modify their behaviour as a function of the changing opportunities and pressures, as they each pursue their own goals, for the location and re-location of employment according to the functional requirements, and as private citizens, as a function of their means and the opportunities. The spatial dynamics can therefore generate and capture the complex effects of housing price dynamics, and also the complex effects of planning regulations on commercial and industrial employment, as well as the effects of changes in the transportation systems.

In Fig. 5 we see the interaction diagram of the different types of actor considered adequate to represent the spatial evolution of a city like Brussels in the 1980s. It has different possible interaction mechanisms between them, which express the need for flows of goods services and people between different locations, and also the pressure of spatial concentration affecting land prices and rents.

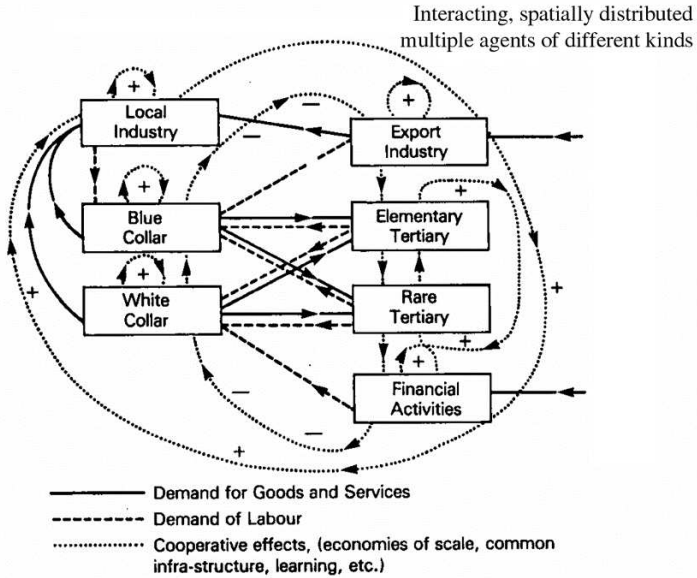


Fig. 5. The interaction diagram that can provide the essential mechanism operating within the city.

The mechanisms above, when run under a scenario of overall growth, spontaneously generate self-consistent urban spatial structure for the 7 types of actor, as well as the corresponding flows of goods, services and people.

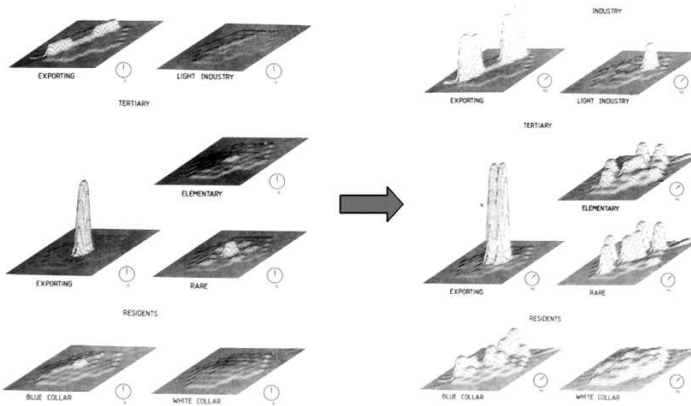


Fig. 6. Emergence of complementary spatial structure of the 7 variables

Such a model can therefore be used to explore the effect on the spatial structure of possible modifications of the transportation system. This could correspond to

plans for new roads, tramways or Metro system. By looking at the changes in structure that follow some intervention, our model can explore the impacts over time of a given action, as actors respond to the new situation, and their changed behaviour affects other actors in turn, creating a complex spatial multiplier (Fig. 4).

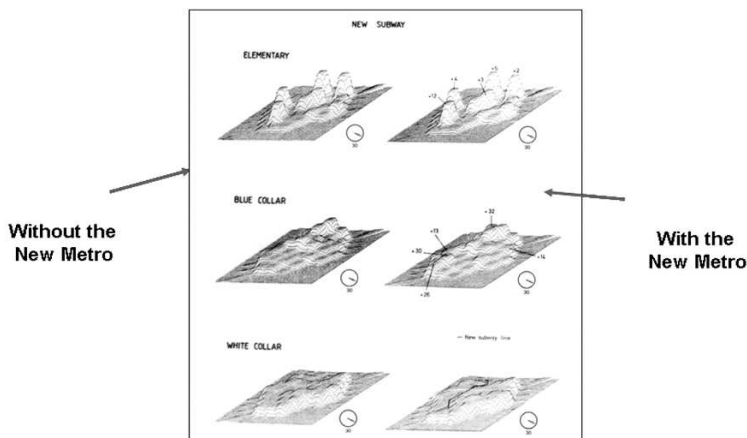


Fig. 7. The distributions of residents and of tertiary activity are shown for simulations with and without a new Metro system.

This allows us to examine the complex effects of the cascading interactions under different possible plans for the Metro: possible routes, locations of stations, train sizes and frequencies. While not pretending that each outcome is a real prediction that is accurate, what matters are the relative differences between simulation outcomes, since these will show the “relative effects” of different routes, of more or less trains etc. Similarly, the model can be used over the longer term to examine some strategic issues such as the effect on decentralisation – centralisation. This surely is one of the major questions that affects any city – will this action influence the existing trends and patterns of migration of jobs and people into the periphery? Clearly, the action of building Metro systems in one that tends to “allow” people to travel to the central part of the city with some ease. By the additional use of “park and ride” car parks at edge of town metro systems it may even encourage the further out-migration of residents from the city, but anchor employment at its centre. The point about this is that any evaluation of the plans for a Metro System need to include not only the projected costs of the system, but also the projected effects on the city. These projected effects cannot be calculated simply from the expected “traffic” that switches from roads to the Metro, but also needs to encompass the spatial changes brought about to the residential, commercial and employment sectors. In particular, it is important to have some idea of the strategic impacts of a transportation scheme, and whether it will tend to accelerate or reverse some basic trends that are running in the system at present.

Despite being developed over a decade ago, models of the kind described above are still not used by decision makers such as regional, urban or local authorities. It seems clear then that such decisions are left to the intuitive judgement of such authorities, acting under advice and pressure of competing lobbies and groups with particular interests. This cannot be a good way to make decisions. In particular, it seems evident that some method is required for estimating the strategic impacts on the growth patterns of the city, effects on house prices, on residential and commercial development, and in turn on future traffic patterns, energy consumption, pollution etc. The framework discussed briefly here is a candidate for this, and therefore deserves a renewal of interest in its development and adoption.

In recent years, the model described above was up-dated and the program re-written for potential applications.

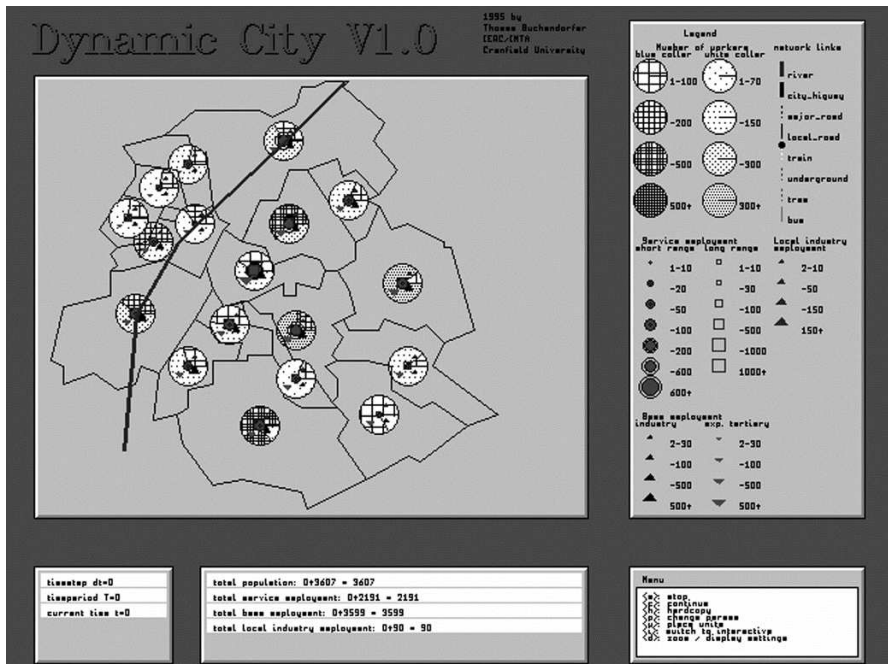


Fig. 8. The updated model of Brussels, developed by T. Buchendorfer and the author.

In further work too, the Asian Development Bank has also commissioned work designed to explore the strategic spatial consequences of different possible transport investment plans in West Bengal. This enabled questions like the spatial distributions of the impacts on poverty to be assessed. This is something that has become a necessary pre-condition for many projects to be sanctioned by international organisations like this, but in fact there appears to be no method of calculating such impacts other than those described above. Some methods have been developed and applied to transport decisions in West Bengal.

4 Manufacturing Evolution

The previous sections demonstrate theoretically how micro-diversity in character space, tentative trials of novel concepts and activities, will lead to emergent objects and systems. However, it is still true that we cannot predict what they will be. Mathematically we can always solve a given set of equations to find the values of the variables for an optimal performance. But we do not know *which* variables will be present, as we do not know what new “concept” may lead to a new structural attractor, and therefore we do not know *which* equations to solve or optimise. The changing patterns of practices and routines that are observed in the evolution of firms and organisations can be looked at in exactly the same way as that of “product” evolution above. We would see a “cladistic diagram” (a diagram showing evolutionary history) showing the history of successive new practices and innovative ideas in an economic sector. It would generate an evolutionary history of both artifacts and the organisational forms that underlie their production (McKelvey 1982, 1994; McCarthy 1995; McCarthy et al. 1997). Let us consider manufacturing organisations in the automobile sector.

Standardisation of Parts	1	TQM sourcing	27
Assembly Time Standards	2	100% inspection sampling	28
Assembly line layout	3	U-Shape layout	29
Reduction of Craft Skills	4	Preventive Maintenance	30
Automation (Machine paced shops)	5	Individual error correction	31
Pull Production System	6	Sequential dependency of workers	32
Reduction of Lot size	7	Line balancing	33
Pull procurement planning	8	Team Policy	34
Operator based machine maintenance	9	Toyota verification of assembly line	35
Quality circles	10	Groups vs. teams	36
Employee innovation prizes	11	Job enrichment	37
job rotation	12	Manufacturing cells	38
large volume production	13	Concurrent engineering	39
mass sub-contracting by sub-bidding	14	ABC Costing	40
exchange of workers with suppliers	15	Excess capacity	41
Training through socialisation	16	Flexible automation of product versions	42
Proactive training programmes	17	Agile automation for different products	43
Product range reduction	18	in-Sourcenge	44
Automation (Machine paced shops)	19	Immigrant workforce	45
Multiple sub-contracting	20	Dedicated automation	46
Quality Systems	21	Division of Labour	47
Quality Philosophy	22	Employees are system tools	48
Open Book Policy with Suppliers	23	employees are system developers	49
Flexible Multifunctional workforce	24	product focus	50
set-up time reduction	25	Parallel processing	51
Kaizen change management	26	Dependence on written rules	52
		Further intensification of labour	53

Fig. 9. 53 Characteristics of manufacturing Organisations

With these characteristics (Fig. 9) as our “dictionary” we can also identify 16 distinct organisational forms:

- Ancient craft system
- Standardised craft system
- Modern craft system
- Neocraft system
- Flexible manufacturing

- Toyota production
- Lean producers
- Agile producers
- Just in time
- Intensive mass producers
- European mass producers
- Modern mass producers
- Pseudo lean producers
- Fordist mass producers
- Large scale producers
- Skilled large scale producers

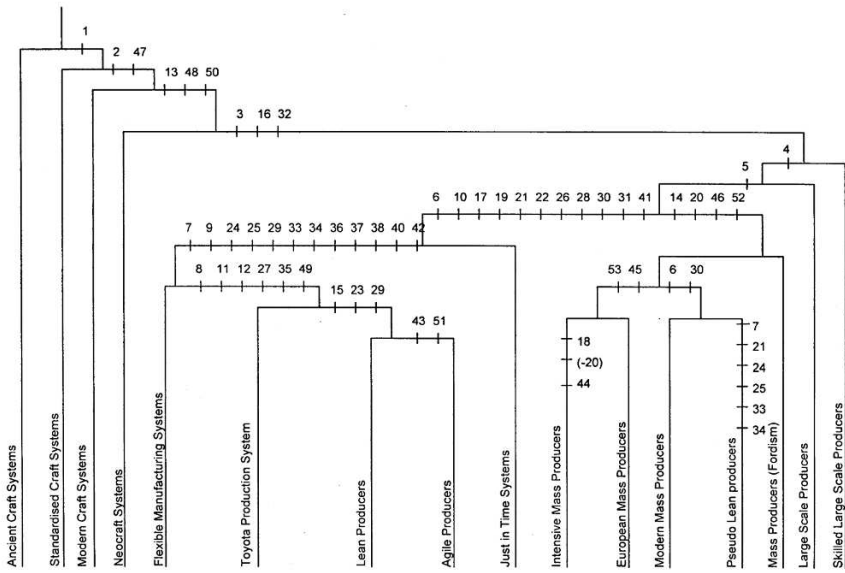


Fig. 10. The cladistic diagram for automobile manufacturing organisational forms (McCarthy et al. 1997)

The evolutionary tree of Fig. 10 can be deduced from cladistic theory, and this shows the probable sequence of events that led to the different possible organisational forms. However, in the spirit of complex systems thinking and that of the formation of networks, we want to consider the synergy or conflict that different *pairs of attributes* actually have. Instead of only considering the different list of characteristic features that constitute the different organisational forms, we also look at the pair-wise interactions between each pair of practices, in order to examine the role of “internal coherence” on the organisational performance. In this “complex systems” approach, a new practice can only invade an organisation if it is not in conflict with the practices that already exist there. In other words, we are

looking at “organisations” not in terms of simply additive features and practices, but as mutually interactive “complexes” of constituent factors.

From a survey of manufacturers (Baldwin et al. 2003) concerning the positive or negative interactions between the different practices, a matrix of pair interaction was constructed allowing us to examine the “reasons” behind the emergent organisational forms, with successful forms arising from positive mutual interactions of constituent practices. This is shown in Fig. 11.

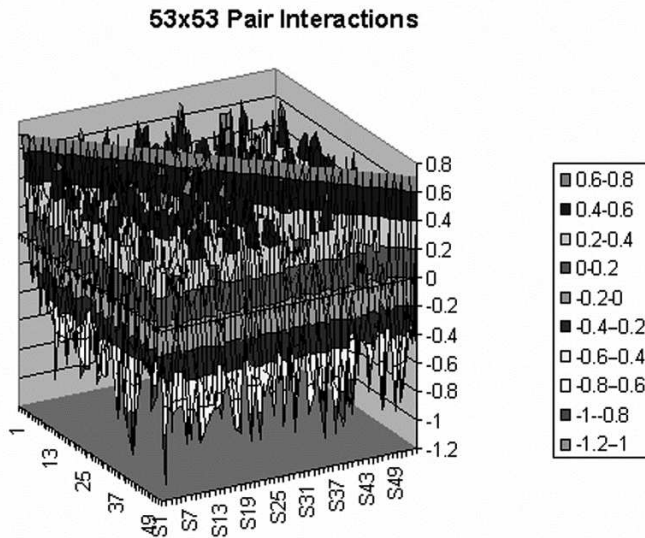


Fig. 11. The 53x53 matrix of pair interactions of the characteristic practices. It allows us to calculate the net attraction or conflict for any new practice depending on which ones are present already.

We have then been able to develop an evolutionary simulation model, in which a manufacturing firm attempts to incorporate successive new practices at some characteristic rate. There is an incredible range of possible structures that can emerge, however, depending simply on the order in which they are tried. But, each time a new practice is adopted within an organisation, it changes the “invadability” or “receptivity” of the organisation for any new innovations in the future. This is true illustration of the “path dependent evolution” that characterises organisational change. Successful evolution is about the “discovery” or “creation” of highly synergetic structures of interacting practices.

In Fig. 11 we see the changing internal structure of a particular organisation as it attempts to incorporate new practices from those available. In the simulation, the number available start from the ancient craft practice on the left, and successively add the further 52 practices on the right. At each moment in time the organisation

can choose from the practices available at that time, and its overall performance is a function of the synergy of the practices that are tried successfully. We see cases where practice 4, for example, is tried several times and simply cannot invade. However, practice 9 is tried early on and fails, but does successfully invade at a later date. The particular emergent attributes and capabilities of the organisation are a function of the particular combination of practices that constitute it.

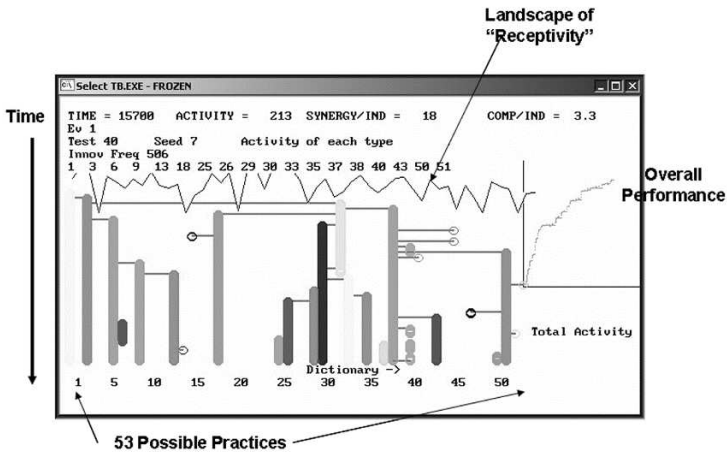


Fig. 12. An evolutionary model tries to “launch” possible innovative practices in a random order. If they invade, they change the “invadability” of the new system.

The model starts off from a craft structure. New practices are chosen randomly from those available at the time and are launched as a small “experimental” value of 5. Sometimes the behaviour declines and disappears, and sometimes it grows and becomes part of the “formal” structure that then conditions which innovative behaviour can invade next.

Different simulations lead to different structures, and there is a very large number of possible “histories”. This demonstrates a key idea in complex systems thinking. The explorations/innovations that are tried out at a given time cannot be logically or rationally deduced because their overall effects cannot be known ahead of time. Therefore, the impossibility of prediction gives the system “choice”. In our simulation we mimic this by using a random number generator to actually choose what to try out, though in reality this would be promoted by someone who believes in this choice, and who will be proved right or wrong by experience, or in this case by our simulation. In real life there will be debate and discussion by different people in favour of one or another choice, and each would cite their own projections about the trade-offs and the overall effect of their choice. However, the actual success that a new practice meets with is pre-determined by the “fitness landscape” resulting from the practices already present and what the emergent attributes and capabilities encounter in the marketplace. But this landscape will be changed if a new practice does successfully invade the

system. The new practice will bring with it its own set of pair interactions, modifying the selection criteria for further change. So, the pattern of what *could* then invade the system (if it were tried) has been changed by what *has already* invaded successfully. This is technically referred to as a “path dependent” process since the future evolutionary pathway is affected by that of the past.

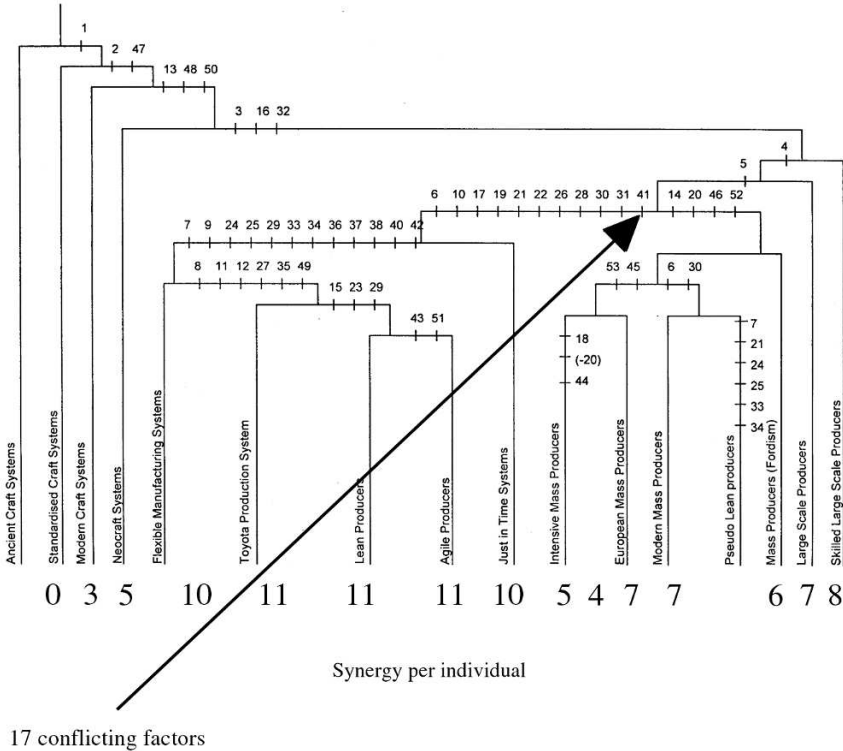


Fig. 13. Knowledge of the pair matrix for the different characteristics allows us to calculate the synergy/individual in the different organisations.

Our results have already shown (Fig. 13) that the evolution through the tree of forms corresponds to a gradual increase in overall “synergy”. That is, the more modern structures related to “lean” and to “agile” organisations contain more “positive” links and less “negative” links per unit than the ancient craft systems and also the mass-producing side of the tree. In future research we shall also see how many different structures could have emerged, and start to reflect on what new practices and innovations may be available today for the future.

Our work also highlights a “problem” with the acceptance of complex systems thinking for operational use. The theory of complex systems tells us that the future is not completely predictable because the system has some internal autonomy and

will undergo path dependent learning. However, this also means that the “present” (existing data) cannot be proven to be a *necessary* outcome of the past – but only, hopefully, a *possible* outcome. So, there are perhaps so many possible structures for organisations to discover and render functional, that the observed organisational structures may be 16 in several hundred that are possible. In traditional science the assumption was that “only the optimal survive”, and therefore that what we observe is an optimal structure with only a few temporary deviations from average. But, selection is effected through the competitive interactions of the other players, and if they are different, catering to a slightly different market, and also sub-optimal at any particular moment, then there is no selection force capable of pruning the burgeoning possibilities to a single, optimal outcome. Complexity tells us that we are freer than we thought, and that the diversity that this freedom allows is the mechanism through which sustainability, adaptability and learning occur.

This picture shows us that evolution is about the discovery and emergence of structural attractors (Allen 2001) that express the natural synergies and conflicts (the non-linearities) of underlying components. Their properties and consequences are difficult to anticipate and therefore require real explorations and experiments, to be going on, based in turn in diversity of beliefs, views and experiences of freely acting individuals.

5 Structural Attractors

There are several important points about these results. The first is that the model above is very simple, and the results very generic. It shows us that for a system of co-evolving agents with underlying microdiversity and idiosyncrasy, then we *automatically* obtain the emergence of structural attractors such as Fig. 11. A structural attractor is the temporary emergence of a particular dynamical system of limited dimensions, from a much larger space of possible dynamical systems and dimensions. These are complex systems of interdependent behaviours whose attributes are on the whole synergetic. They have better performance than any single, pure homogeneous behaviour, but are less diverse than if all “possible” behaviours were present. In other words, they show how an evolved entity will not have “all possible characteristics” but will have some that fit together synergetically, and allow it to succeed in the context that it inhabits. They correspond to the emergence of hypercycles in the work of Eigen and Schuster (1979), but recognise the importance of emergent collective attributes and dimensions. The structural attractor (or complex system) that emerges results from the particular history of search and accident that has occurred and is characteristic of the particular patterns positive and negative interactions of the components that comprise it. In other words, a structural attractor is the emergence of a set of interacting factors that have mutually supportive, complementary attributes.

What are the implications of these structural attractors:

- search carried out by the “error-making” diffusion in character space leads to vastly increased performance of the final object. Instead of a homogeneous system, characterised by intense internal competition and low symbiosis, the development of the system leads to a much higher performance, and one that decreases internal competition and increases synergy.
- the whole process leads to the evolution of a complex, a “community” of agents whose activities, whatever they are, have effects that feed back positively on themselves and the others present. It is an emergent “team” or “community” in which positive interactions are greater than the negative ones.
- The diversity, dimensionality and attribute space occupied by the final complex is much greater than the initial homogeneous starting structure of a single population. However, it is much less than the diversity, dimensionality and attribute spaces that all possible populations would have brought to the system. The structural attractor therefore represents a reduced set of activities from all those possible in principle. It reflects the “*discovery*” of a subset of agents whose attributes and dimensions have properties that provide positive feedback. This is different from a classical dynamic attractor that refers to the long-term trajectory traced by the given set of variables. Here, our structural attractor concerns the *emergence* of variables, dimensions and attribute sets that not only coexist but actually are synergetic.
- a successful and sustainable evolutionary system will clearly be one in which there is freedom and encouragement for the exploratory search process in behaviour space. Sustainability in other words results from the existence of a capacity to explore and change. This process leads to a highly co-operative system, where the competition per individual is low, but where loops of positive feedback and synergy are high. In other words, the free evolution of the different populations, each seeking its own growth, leads to a system that is more co-operative than competitive. The vision of a modern, free market economy leading to, and requiring a cut-throat society where selfish competitiveness dominates, is shown to be false, at least in this simple case.

The most important point really is the generality of the model presented above. Clearly, this situation characterizes almost any group of humans: families, companies, communities etc., but only if the exploratory learning is permitted will the evolutionary emergence of structural attractors be possible. If we think of an artifact, some product resulting from a design process, then there is also a parallel with the emergent structural attractor. A successful product or organisation is one in which the “bundling” of its different components creates emergent attributes and capabilities that assure the resources for its production and maintenance. However, the complication is that the emergent attributes and capabilities are not simply an additive effect of the components. If a change is made in the design of one component it will have multi-dimensional consequences for the emergent properties in different attribute spaces. Some may be made better and some worse. Our emergent structural attractor is therefore relevant to understanding what successful products and organisations are and how they are obtained. Clearly, a successful product is one that has attributes that are in synergy, and which lead to a

high average performance. From all the possible designs and modifications we seek a structural attractor that has dimensions and attributes that work well together.

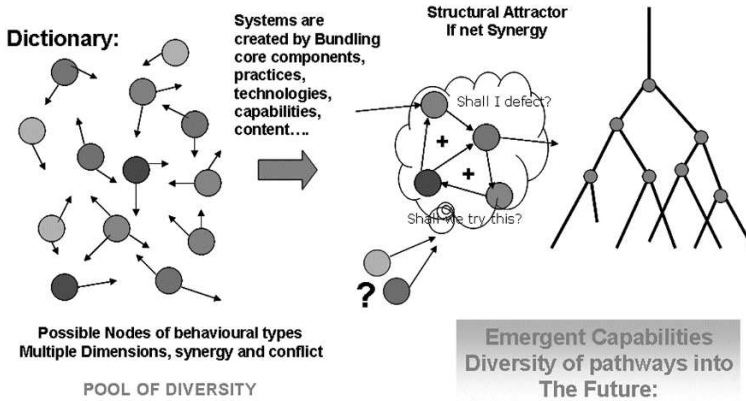


Fig. 14. On the left we have “dictionary” of possible core concepts, practices or ideas. These are “bundled” on the right and if the different elements have synergy then the structure is successful.

The structural evolution of complex systems is as shown in Fig. 14 how explorations and perturbations lead to attempts to suggest modifications, and these lead sometimes to new “concepts” and structural attractors that have emergent properties. The history of any particular product sector can then be seen as an evolutionary tree, with new types emerging and old types disappearing. But in fact, the evolution of “products” is merely an aspect of the larger system of organisations and of consumer lifestyles that also follow a similar, linked pattern of multiple co-evolution.

6 Conclusions

The conclusion of this contribution is that the new ideas emerging from complex systems thinking offer us a new basis for understanding and living in the real world. Since the possibility of structural change, learning and innovation are considered, these kind of models provide a new basis for policy exploration, particularly with respect to issues of “sustainable” development. In these the “bio-physical” part of the system (the hydrology, soils, vegetation, ecology, physical infrastructure etc.) is linked dynamically to the “human” part of the system that is driving the exploitation of resources, both natural and human.

These developments underline the fact that these models should not be thought of as only of “academic” interest. Nor are they just biological or chemical meta-

phors. The fundamental points that have been made concern the scientific basis of understanding. Understanding is achieved in a trade-off between simplicity and realism. The whole question is whether or not a simple enough description can be found which is still sufficiently realistic to be useful. In the past, the desire for tractability has led to the use of very strong assumptions such as that of "equilibrium", which is necessary prerequisite for a normal "cost/benefit" analysis of a decision. It is our contention here, that such methods are incorrect – although possibly better than nothing. The new methods presented here are still not used operationally, which means that any strategic aims that are involved in the decision to invest in new transportation systems are really based on the personal intuition of the people involved. Of course, these can be correct, but in general it would be good to be able to provide better information about the probable consequences of such schemes.

The history of a successful society within a region is largely a tale of increasing cooperation and complementarity, not competition. An economy is a "complex" of different activities that to some extent "fit together" and need each other. Competition for customers, space, or for natural resources is only one aspect of reality. Others are of familiar suppliers and markets, local skill development and specialization, co-evolution of activities to each other, networks of information flows and solidarities, that lead to a collective generation and shaping of exchanges and discourse within the system. Evolution is not about a single type of behaviour "winning", through its superior performance, but rather by increasing diversity and complexity. The models we propose are therefore ones that can help us deal with the overall, integrated effects of the coupled decisions of multiple actors, and allowing us better insight into the consequences of possible policies and actions.

References

- Allen PM (1976) Evolution, Population Dynamics and Stability. *Proc Nat Acad Sci* **73**: 3: 665-668
- Allen PM (1983) Self-Organisation and Evolution in Urban Systems. *Cities and Regions as Nonlinear Decision Systems*, AAAS Selected Symposia **77**: 29-62, Ed Crosby Westview press, Boulder
- Allen PM (1990) Why the Future is not what it was. *Futures* **22**: 6: 555-569
- Allen PM (1997a) Cities and Regions as Evolutionary Complex Systems. *Geographical Systems* **4**: 103-130
- Allen PM (1997b) *Cities and Regions as Self-Organizing Systems: Models of Complexity*. Gordon and Breach, Environmental Problems and Social Dynamics Series, Amsterdam
- Allen PM (2000) Report to the Asian Development Bank. A Regional Poverty Impact Assessment for Transport Infrastructure Investments In West Bengal. Prepared in conjunction with the Report by Halcrow Fox, Vineyard House, 44 Brook Green London W6 7BY. North-South Corridor Development Project in West Bengal
- Allen PM (2001) A Complex Systems Approach to Learning in Adaptive Networks. *International Journal of Innovation Management* **5**: 2: 149-180

- Allen PM, Sanglier M, Engelen G, Boon F (1985) Towards a New Synthesis in the Modelling of Evolving Complex Systems. *Environment and Planning B, Planning and Design: Special Issue*
- Allen PM, Strathern M, Baldwin JS (2005) Understanding Social and Economic Systems as Evolutionary Complex Systems. In: McDaniel RR, Driebe DJ (eds) *Uncertainty and Surprise in Complex Systems: Questions on Working with the Unexpected*. Springer, Berlin
- Baldwin JS, Allen PM, Winder B, Ridgway K (2003) Simulating the Cladistic Evolution of Manufacturing. *Journal of Innovation Management, Policy and Practice* **5**: 2-3: 144-156
- Buchendorfer T (1998) *Dynamic Models of Urban Systems*. PhD Thesis, Cranfield University
- Eigen M, Schuster P (1979) *The Hypercycle*. Springer, Berlin
- McCarthy I (1995) Manufacturing Classifications: lessons from organisational systematics and biological taxonomy. *Journal of Manufacturing and Technology Management-Integrated Manufacturing systems* **6**: 6: 37-49
- McCarthy I, Leseure M, Ridgway K, Fieller N (1997) Building a Manufacturing Cladogram. *International Journal of Technology Management* **13**: 3: 2269-2296
- McKelvey B (1982) *Organizational Systematics*. University of California Press CA
- McKelvey B (1994) Evolution and Organizational Science. In: Baum J, Singh J (eds) *Evolutionary Dynamics of Organizations*. Oxford University press, pp 314-326
- Prigogine I (1980) *From Being to Becoming: Time and Complexity in Physical Sciences*. W.H. Freeman, San Francisco

Ontogeny and Ontology in Complex Systems Modeling

Claes Andersson

Department of Physical Resource Theory, Chalmers University of Technology,
Göteborg, Sweden

Abstract

In this paper the ontogeny of complex systems models is discussed: the historical aspect of model ontology. The theoretical framework that is applied is complex systems theory and more specifically evolution and dynamical hierarchies. Some issues relating to the role and applicability of complex systems models are also discussed.

1 Introduction

The strength of human control over urban systems declines dramatically the longer scales of time and space that are observed; most macroscopic features of urban systems are neither designed nor subject to efficient selection, and they are thereby caused by human action but hardly intended. In other words, changes to the state of the urban system are introduced with only limited foresight, a weaker form of rationality which Simon termed *bounded rationality* (Simon 1969). Bounded rationality is the inescapable consequence of the impossibility of predicting the future evolution of complex systems. Furthermore, capital investments in the form of land improvements and machinery (both collective such as e.g. infrastructure and “private” on the level of firms, households, et cetera) are costly to change and this makes system states highly inert. These two factors combined: bounded rationality and high inertia severely limits the applicability here of economic models based on neoclassical assumptions of rationality and general equilibrium. Instead, urban growth is increasingly viewed as the application of stationary or nearly stationary micro mechanisms that dynamically introduce small alterations to the system. This is the perspective employed in most simulation models of today such as cellular automata and agent-based models (see e.g. Andersson et al. 2002; Clarke et al. 1997; Couclelis 1985; Sanders et al. 1997; Tobler 1979; Torrens and O’Sullivan 2001; White and Engelen 1993), and also pattern, cluster and morphology models (e.g. Batty and Longley 1994; Gabaix 1999; Makse et al. 1998, 1995; Schweitzer and Steinbrink 1998).

The realization that urban systems are inert and are changed with the guidance of only bounded rationality is of course of an old date, but the viability of a scientific approach based on these properties hinges to not a small extent on computa-

tion. Such a “complex systems approach” to urban modeling, just as in parallel developments in a range of other fields, was impossible to realize before computers became widely available. Previous models of urban systems such as central place theory, monocentric city models, systems of cities model, entropy models, spatial interaction models and industrial location theory were formed around the tools that were available at the time (Alonso 1964; Christaller 1933; Henderson 1974; Hotelling 1929; Isard 1960; Weber 1997; Wilson 1970), yet some are still today formed with an adherence to this formal framework such as the “new economic geography” (Fujita et al. 1999). Without the aid of computers, complex systems aspects had to be assumed away or circumvented and mainstream economic theory appears to have formed and hardened around this set of analytic tools which were basically imported from classical mechanics. This has made it hard for already highly formalized areas to incorporate new theory based on different sets of tools, see for example a discussion by Saviotti on this topic (Saviotti 1996, pp. 22-28). On the other hand, less formalized areas of social science have had more flexibility in this respect.

Still, this flexibility of complex systems models has to be viewed as a freedom that comes with responsibility: it is easy to incorporate too much, which threatens not only to make the models unrealistic but also highly opaque. As Nelson and Winter remarked in relation to evolutionary models: “if reality could be ‘copied’ into a computer program, that approach might be productive - but it cannot, and it is not” (Nelson and Winter 1982). This remark is indeed from the early days of computer modeling, but the reasons behind it are not primarily related to computational complexity but to the compound nature of assumptions - making ten slightly poor assumptions might be worse than making two bad ones - and to the problems with producing exhaustive lists of what to include in a model and how to include it.

2 Dynamical Hierarchies

Talking about a micro-macro dichotomy in complex systems is often misleading. Except for in very simple physical systems, new dynamics appear several times as one observes systems over longer scales of time and space. Instead, a hierarchical perspective has been proposed for complex systems in a variety of fields, including economics, biology and engineering, see e.g. Gould (2002), Nelson and Winter (1982) and Simon (1969). Objects join together in new objects that take part in new dynamics, and this can take place recursively since the formation of a new object effectively hides the complexity of the smaller-scale dynamics. Many of the properties of new objects are emergent, which means that they are not applicable to the constituent parts of the object. For example, that one can edit texts using a computer is a property of the computer as a system - there is nothing about the components of a computer in isolation that suggest that this can be done. This encapsulation of smaller dynamics introduces redundancy to the system: redundancy

that can be used for forming simpler models where much complexity can be ignored.

When a new object is formed the complexity that it houses becomes unimportant on the scale on which the object interacts with other objects. The only importance of the smaller scale that remains is the limitations that this smaller dynamics puts on the interface of the object. The practical implication of this is that whichever mechanism that is responsible for the formation of objects (usually intelligence or evolution) can then start over again from scratch. For instance, organisms are made from organs, which are made from tissues, which are made from cells. With such an organization, the amount of information required to direct the ontogeny is dramatically lower than had it been if every cell, or worse, every molecule, had to be positioned in the body. In fact we can conclude that even a cell would be impossible to conceive of without such a hierarchical organization.

There are two important causal directions in such hierarchies that both give rise to important sets of questions. These sets of questions are at the same time related and unrelated in important ways. First, there is vertical causation - the issue of origins and consistency of the scientific system. If the dynamics on one level is explained in terms of objects on this level, then this explanation is valid only to the extent that the assumptions made about the objects are consistent with knowledge about how the more fundamental levels work. For example, biological evolutionary theory provides a scientifically tenable causal link between the mechanistic physical world and the adaptive biological world. This is in contrast with other early theories with the same range, such as Lamarckism. Darwin assumed myopic change which is and was consistent with what is known about the smaller scales of nature. Lamarckism on the other hand requires mechanisms of change that are impossible to reconcile with physics and chemistry. It is important to emphasize that such a causal link in no way implies understanding of phenomena - we learn very little about human behavior from studying cells since emergent properties appear on higher levels. We would however not be entirely comfortable if the behavior of cells would be unknown and impossible to study. The point of "vertical inquiries" is not primarily to improve macroscopic models; in fact, discerning the workings of a lower level in the hierarchy does not necessarily (or even usually) change existing more macroscopic models but it improves the consistency of the scientific system¹. Moreover, such inquiries lack logical end points. Hierarchical objects are like Russian Matryoshka nesting dolls: every time we pick an object apart to study its innards we are faced with a new set of objects begging to be explicated.

Secondly, it is the question of horizontal causation: the behavior of a system viewed at some specific level. We may for instance fruitfully model a social system on the basis of agents behaving in such and such ways. For such inquiries, most questions concerning vertical causal links up and down the hierarchy can be deferred. As mentioned, objects can be defined in terms of interfaces rather than in terms of their internal mechanisms. This is of course fortunate since we would

¹ For example, classical mechanics remains valid alongside mechanics and relativity theory. What has changed and improved is not the content of the older classical theory but the picture of its range.

otherwise be unable to think about complex systems and much less to model them. This article will almost exclusively concern this second type of inquiries about complex systems.

3 Ontogeny

When we construct models of urban systems or any other social system, we have to make use of a “clean slate” at some point: a level where we treat objects according to their interface without explicit reference to their internal dynamics. The methodology proposed here is to use a hierarchical model architecture - to realize that there exists no universal micro and macro dichotomy but in fact a whole hierarchy of levels on which models can be formulated. What we need to do is to decide how to treat every single object in the world: a task that fortunately sounds more daunting than it is. In the majority of cases the explicit inclusion of the object is so absurd that it never even crosses our mind. For example, in a model of urban growth, we do not explicitly reject plate tectonics and the citric acid cycle. However, as we get near the scale on which the dynamics of the model exists, the decisions become harder and harder and deliberate decisions have to be made. Note that these decisions may very well lead to a model without any explicit objects: there are many fruitful models that include no explicit objects or interactions whatsoever. I will here use one absurd example and one relevant example in tandem for each situation to illustrate the point. Large/slow objects are typically treated as constant, e.g. we assume that continents remain where they are but on a more relevant scale we might also assume that institutions remain fixed for the time frame of the model run. Too small/fast objects are either treated statistically, as non-existing or as internal to objects of which we concern ourselves only with the interface. E.g. we treat bodies of water and air as behaving in predictable ways or that fluctuations in their behavior will sufficiently even out over the duration of the run. Again, on more relevant scales we might treat the flows of people and goods as instantaneous on the time scale that we are dealing with and that human decisions can be modeled stochastically from the basis of some identified important factors that influence decisions. The stochastic treatment of small but relevant degrees of freedom reflects a lack of representation and information about the particularities of the smaller parts of the hierarchy: we can not leave this dynamics out of the model and we can not include it specifically so we treat it in some aggregate way. In formal models, a lack of knowledge about the micro scale is often dealt with as constrained maximization of entropy using a Lagrangian formalism and in simulation models it is generally captured with a pseudo random number generator. The citric acid cycle we treat as internal to persons and we trust the mechanisms of homeostasis² to allow people to be sufficiently persistent in time to

² Homeostasis is a term used in biology to refer to mechanisms to maintain a stable inner-environment. For instance, the mechanisms that maintain a constant body temperature of mammals and birds is a homeostasis mechanism. More generally, the term is used to de-

perform the actions that we model on a more aggregated scale. On the more relevant scale in this case we might treat the internal dynamics of firms as hidden behind an interface of the firm. There is of course no prescribed universal methodology for doing this, but the fact that this is in fact necessary to do must be before us as we construct models.

We thus start from a very simple and general model and introduce content based on how important we think things are for the problem in question and how they interact with the phenomena that we wish to study. For instance, in the case of urban systems, we may deem that the size of roads are important but the individual cars that travel there are not. In such a fashion we determine what to treat as constant, what to treat statistically, what to include explicitly and what to ignore altogether. Having done this, we have typically narrowed the model down to a certain scale and position in time and space and what remains is parameter values. The structure and time-ordering of model construction is however of great relevance. We generally want to “draw the line” somewhere and this is sometimes easy and sometimes very hard. It is easy if we want to for example approximate an infinitely differentiable function around a value with a polynomial: we make a series expansion and include some number n of terms. However, we always include the first n terms, not any n terms. The same logic applies also to other models although the problem might be considerably less well defined, see Fig. 1.

For a series expansion there is a unique way to accomplish this, but in the case of models of complex phenomena the situation is much more intricate. The basic logic still remains however, detailed behavior makes sense only in the context of *that of which it is a detail*. There is therefore a highly contingent component not only of the output of complex systems models, but also of their construction. A series expansion is highly sterile, so perhaps a parallel to embryonic growth is more suited: ontogeny, the construction of an organism from its genetical description, reflects the contingent nature of the formation of something specific out of something general. As can be seen in Fig. 2 embryos of different vertebrates are very similar during early development and only with subsequent development is the original body plan refined to have the adaptations that we connect with specific species.

Going from the general towards the specific is easier and more robust than going the other way around or to create a complicated model directly. Doing so demands that the original complicated setup is exhaustive on the scale on which it is formulated, and this is very hard to ensure in most cases. Starting out with a general model and adding more and more components to it on the other hand allows one to evaluate a) whether the additions bought any new predictive or explanatory power and b) if the less specific predictions and explanations generated by the more general parent model still remain. This also allows the investigation of “when” model behavior appears, that is, it allows the investigation of what the minimal model is for capturing some behavior. At each step it should be ensured

note mechanisms that maintain a stability and persistence of objects – that make their internal environment resistant to exogenous forces.

that additions are made with as little restriction as possible to future extensions; that is, one should strive to facilitate future relaxation of made assumptions.

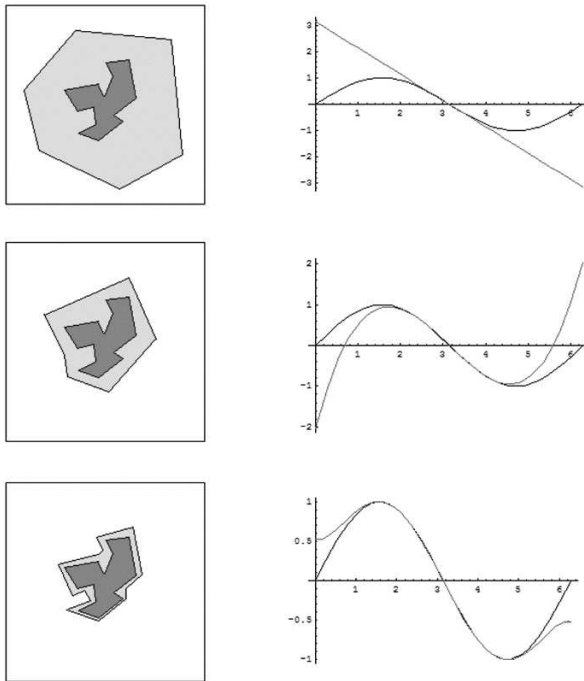


Fig. 1. To the right is shown a series expansion of $\sin x$ around $x = \pi$ and a $\sin x$ curve. As more and more terms in the expansion are included (1, 3 and 5 terms) the polynomial approximates the sine curve in the neighborhood better and better. It is easily verified that the inclusion of a new term makes sense only with the inclusion of all lower-order terms. In the left column a conceptualization of a more abstract modeling case is shown. In the top figure, a very general model has the sought-for state in its range but it is also consistent with a rather wide range of other states. In the figures below, the model is made more and more specific in a fashion analogous with the series expansion example, and the range still includes the true state but fewer and fewer other states.



Fig. 2. Shown in this figure are embryos of fish, chicken, dog, human and lizard. With knowledge only about the post natal forms of these species it is very hard, if not impossible, to tell which is which. The construction of the body, just as the evolutionary construction of

the genome, necessarily proceeds from the simple and undifferentiated towards the more complex and specific.

4 The Network Model

The economy consists of interacting agents that are immobile on shorter time scales. Thus, one can speak of a geographical network of activities - a network that is very large and that changes over time. These two properties qualify complex network models as potential models for the analysis of the economic geography. A series of simple such models have been reported previously where some aggregated stationary properties of the economic geography were reproduced (Andersson et al. 2003, 2005a, 2005b). These models explain only aggregate properties of urban systems, and it is clear that they are thereby susceptible to the common critique against complex systems models that they are too general to be of much value. In this paper some possible roles of such models is discussed, and how the presented ontogenetic principles have influenced the design of the mentioned network models.

Table 1. Incremental model construction

Model component added	Micro mechanisms added	Prediction
Multiplicative growth	Statistical assumption that if we know nothing except the size of cities, the probability of a growth event taking place in a city must be assumed to be its relative size.	Power law distributed city sizes
Network with multiplicative and uniform growth	Interactions and correlations of growth. Multiplicative growth logic applied to values of land lots rather than sizes of cities. The land market is modeled as a fast process and approximated as being close to equilibrium.	Power law distributed land values per unit area rather than city sizes.
Spatial cellular distribution with nearest-neighbor infrastructure availability sub-model	Edge growth and new cluster formation	Power law distributed city sizes appear alongside power law distributed land values. Fractal properties of the clusters are also in agreement with empirical observations.

A complex network representation of the geography of human activities is indeed firmly in the evolutionary and historical perspective on urban evolution. Still, the properties that we have studied so far are in fact not historically path dependent, they are stationary properties of the urban system as it grows. In the papers

that we have published on this model we have compared statistical distributions in simulated systems with a geographic data set over Sweden from year 2000. These statistical distributions appear robustly and are smoothly dependent on model parameter in any run of the computer model and many are in fact analytically tractable with rather innocuous assumptions. Needless to say, the predictive value of such a model is not high: all or nearly all micro states reachable by the process at some parameter values correspond to identical properties. In fact, the model is almost devoid of content. Then what is the value of such a model, or other models like it? As stated by Gabaix, the rank size rule and other global and robust properties of systems can be used as criteria for admissibility of models (Gabaix 1999). Furthermore, provided that such properties are robust for minor alterations of the model, such models can be used as kernels for more specific models.

The complex network models of urban growth were formed in the fashion outlined here and the steps involved are outlined in table I. The fundamental logic of this model is the same as that of Simon's model (Simon 1955): that multiplicative growth can be a suitable starting point if we have a system where we have parallel growth of similar components. The similar components in this case is firms and such, and even if firms are of different size, their growth rates are generally taken to be proportional to their size. Therefore it does not really matter how many firms that are located in an area, their aggregated absolute rate of growth must, in the absence of more detailed knowledge, be assumed to be proportional to the sum of their sizes. This assumption stems originally from Gibrat and is usually termed either Gibrat's Law or the Law of Proportionate Effect, for a review see e.g. (Sutton 2005). This is also how Gibrat's Law extends to urban growth. The approximate size distribution of city sizes (at least for the upper tail) is a power law in most parts of the world, something that originally was identified by Zipf (1949).

Measurements on data over all land values in Sweden in year 2000 aggregated to 100x100 meter cells indicated that also land values are power law distributed (Andersson et al. 2003). It was therefore decided to apply Gibrat's Law to this smaller scale. This could in principle be accomplished by simply re-labeling the cities of previous multiplicative models as land lots. The question was, can something additional be explained by extending the model slightly? One of the main reasons why previous multiplicative growth models and the insights they yield have not been very useful for geographic modeling is that these models are not geographical: there are no interactions and thus there is no relative locations of the cities. Interactions are of central importance to any explicitly geographic model and since growing scale-free networks combine multiplicative growth with interactions in a general way, it was decided to use such a network as a starting point (Barabási and Albert 1999). Growing network models, usually referred to as complex networks, are also open ended because just like in a cellular automaton the update rules can be elaborated to an arbitrary extent.

To be able to map the growth of land values to the network model, a mapping between land value and node degrees had to be found. First of all, we take nodes to be fixed size and non-overlapping land lots. The connection between node degree and land value was then accomplished by assuming the land market to be fast in relation to the time scale of urban growth, and thereby that land values can be

expected to on average be found near an equilibrium. This allows us to use the left-over principle from urban economics, which states that all profit goes to the land owner. In fact, we do not need to assume this to be strictly true, in fact it suffices to assume that some proportion of profits are paid as rent. Complex networks are automata that grow by the addition of connections between nodes through an update rule: if we say that profit is generated by trade - the meeting of some demand by some supply - and if we furthermore do not specify which of the end points that is the buyer and which is the seller, then we can say that the addition of a connection between two nodes in an abstract way represents a new trade connection in the system that will increase the profit and thus the rent of both end-point nodes equally. Since the nodes are taken to be fixed-size cells without overlap, this translates to an increase in rent per unit area in these nodes. We furthermore assume that connections are of unit size and thus we achieve a linear proportionality connection between the count of node connection endpoints (their degree) and land values. In addition to multiplicative growth we also had to introduce an additional type of growth that has a uniform per-node probability of selecting a node to be an end-point of a connection. It appears that it would be harder to argue for the absence of such a component to growth than for its presence: there are many non-multiplicative factors present. Location choice rarely happens on the scale of cells on the size order of 100x100 meter in any case - once a suitable general area has been found other factors outside the model will typically narrow down the search, e.g. the exact location of lots with required land improvements and infrastructure availability, the location of lots that happen to be vacant or have vacant floor space and so on are not multiplicative. Thus, we have something that is a linear combination between a scale free and a random network where the proportion between the type of growth is determined by a parameter. The analysis of this model shows that such a model is consistent with empirical observation on the macro level, and to the extent that the assumptions are sound the approximate power law distribution of per-unit-area land values can be ascribed for with this model; for the model formulation and analysis see Andersson et al. (2003). It should be noted that the model so far in this discussion is not explicitly geographical, there is simply a collection of land lots that interact freely without transportation costs and without any internal structure.

Using geographical cluster analysis on the empirical data set, it was found that the aggregated land value of contiguous clusters of land values above some threshold also conformed well to the rank-size rule. Thus, there are regularities on more than one level of aggregation and it was natural to investigate whether some relatively simple geographical extension of the network model could make it consistent with several observed regularities at the same time. To do this, we embedded the network in a cellular space and introduced spatial structure in two ways: local availability of infrastructure and cost of transportation. Local availability of infrastructure is taken to be binary: either a lot is connected to the infrastructure network or it is not and if it is we make a crude Euclidean approximation of the cost at which it may trade with other lots in the system. This approximation is clearly too crude to ascribe for detailed structure of the system but might not be for the large scale regularities that we wanted to investigate. The model for avail-

ability of infrastructure is a purely local rule that assigns cells with a probability of being connected: unit, medium and minimal. Cells that are already developed (non-zero degree of the cell-node) have already proven that they are connected, cells that are undeveloped but are neighbors of developed cells have medium probability of being connected and finally undeveloped and isolated nodes have a minimal likelihood of being connected. These likelihoods are parameters of the model that can be estimated from data. The reason why they might still be connected in the latter case is that some are passed by the roads that cross the hinterland. Such roads are thus treated probabilistically and not explicitly.

With this extension of the model, the previous behavior of producing approximately power law distributed land values was retained and in addition, the node degrees were spatially arranged in a fashion that is consistent with empirical observation on the level of clusters: aggregated cluster values, the relation between cluster areas and perimeters and the approximate relationship between the area and aggregated value of clusters, for the formulation and analysis of this model see Andersson et al. (2005a, 2005b).

There are at least two types of extensions possible to the model: the structure of the nodes and the selection of nodes for growth. In fact, it is possible to analyze in terms of the model what types of growth rules that are permissible if the basic power law properties are to be retained: the probability of selecting a node must be asymptotically multiplicative. To this end, the expected rate of growth of the nodes can be observed, which is termed the “node fitness” (Andersson et al. 2005). In the full version of the model, the expected growth rate of a node is

$$E[\Delta x_i] = \frac{\eta_i x_i}{\sum_k \eta_k x_k} + \zeta_i \quad (1)$$

where i is an index identifying the node and x_i is the degree of the node. The fitness of a node i is η_i and it measures the deviation from pure multiplicative growth rate of nodes. The growth rate is separated into a multiplicative component and “the rest”, which is denoted ζ_i . Since the rate of growth must be asymptotically multiplicative, it means that as long as ζ_i and η_i are not dependent on x_i in a way such that multiplicative growth is dominated, the distribution of node degrees will be near a power law. This also tells us that power laws in urban systems are, to the extent that the present model is sound, not a result of the agency of agglomeration economies, economies of scale, transportation costs and other often cited forces that affect location choice - it is a result of the lack of agency of these forces on average. This would indicate that such effects are more likely to responsible for deviations from a power law distribution than for adherence to it. It must be noted that this does not suggest that such factors are unimportant in any general sense, they are highly important for the structure of urban systems on a higher resolution but may not be for large scale regularities. In fact, the dependence on parameters and functional form of the function that determines the decrease of connection formation with distance is very weak. The only really important tran-

sportation factor that is required to bring about realistic cluster dynamics is in fact the much more dramatic effect of binary infrastructure availability.

This indicates that the model is highly open ended to further extensions by which the analogue of higher order terms in the series expansion example in Fig. 1 are added to increase the specificity of the model. Specificity means almost inevitably that path dependence becomes important: this happens as soon as we are observing anything other than statistical properties on the macro scale. When this happens, the applicable methods of analysis change fundamentally.

5 Discussion

Over the last decades the direct treatment of complex dynamics in models of urban systems has significantly broadened the range of phenomena that can be studied. Simulation offers a middle ground between discursive theorizing and formal mathematical modeling; apart from providing an injection of ideas to the former two it also constitutes a modeling paradigm in itself, for an overview see e.g. Benenson and Torrens (2004). Rather than viewing system states as equilibria between various economic, technological and social forces, these forces now enter as incremental change mechanism on several scales from the microscopic to the global. Assuming this perspective, more realistic models can be built where the contingent aspect of these systems can be incorporated. The absence of a unique equilibrium means that the history of the system becomes important. States of such systems, for instance the set of cell states in a cellular automaton, can not be explained in the present - they make sense only in the light of the history of the system. This realization has important implications for how we ought to understand economic, social and technical systems and although largely forgotten (Hodgson 2001), this was clearly realized long ago; perhaps earliest and most clearly stated by Friedrich Nietzsche who in the “Genealogy of Morals” wrote:

Perhaps there is no more pregnant principle for any kind of history than the following, which, difficult though it is to master, should none the less be mastered in every detail. - The origin of the existence of a thing and its final utility, its practical application and incorporation in a system of ends, are *toto cœlo* opposed to each other - everything, anything, which exists and which prevails anywhere, will always be put to new purposes by a force superior to itself, will be commandeered afresh, will be turned and transformed to new uses; all “happening” in the organic world consists of overpowering and dominating, and again all overpowering and domination is a new interpretation and adjustment, which must necessarily obscure or absolutely extinguish the subsisting “meaning” and “end”. The most perfect comprehension of the utility of any physiological organ (or also of a legal institution, social custom, political habit, form in art or in religious worship) does not for a minute imply any simultaneous comprehension of its origin: this may seem uncomfortable and unpalatable to the older men, - for it has been the immemorial belief that understanding the final cause or the utility of a thing, a form, an institution, means also understanding the reason for its origin: to give an example of this logic: the eye was made to see, the hand was made to grasp. (Nietzsche 1887)

Indeed, it is obvious that we can not understand why a city looks the way it does, produces what it does or is situated where it is without a reference to its history. Consequently, urban growth models today are invariably evolutionary in the sense that there is little concern for notions such as optimality, rationality and equilibrium. It can be said that a system is evolutionary in a stricter sense if it has adaptive features that can not be ascribed to processes that are internal to the objects. Internal adaptation is guided by foresight and if there is adaptation where the agency of foresight is highly limited, there must be an external adaptive process at play and the only known such process is evolution by natural selection³. Limitations to internal adaptivity include large scales and complexity - the bounds on rationality. The case for evolution in economics has been argued since the 19th century but the perspective has seen a recent upswing over the last decades, see e.g. Nelson and Winter (1982), Saviotti (1996). However, many of the features of cities are not adaptive but quite the opposite. On the level of cities it appears that neither foresight nor selection has much agency. Congestion, fragmentation of biotopes and farmland, pollution and so on are definitely features that are neither designed nor subject to diminishing by selection to any important extent.

The economic geography appears to a great extent to be a collection of externalities (often negative) of the actions of agents on the smaller scale that are guided by bounded rationality. In this sense the urban system in itself is perhaps best likened to an ecological system in which an evolutionary dynamics takes place, and there is in fact nothing that guarantees that such a system evolves towards states of higher utility for their inhabitants. There are a number of factors that can make people act in ways that lower their collective utility. One important factor is that there is a friction between the long run and the short run. Negative externalities often mount slowly over time and the micro actions that would mitigate them might carry with them short run disadvantages to adopters. Absent global coordination, such a situation leads to sub-optimal patterns of behavior, which is analyzed in game theory as the “tragedy of the commons” game (Hardin 1960). Another important phenomenon is lock-in effects. Even if a remedy to a problem is identified, this often happens at a late point where a remedy would be extremely costly and disruptive to implement. This means that the system is irreversible and non-ergodic: it tends to enter parts of state space from which it becomes increasingly unlikely to escape. Lock-ins in economical systems have been studied by several authors over the years, see e.g. Arthur (1994) and Lane (2002). This can happen as a result of increasing returns but also as an effect of specialization. It must also be realized that there is an important “Red Queen” component of societal evolution⁴. To illustrate the reality of such situations, consider a simple priso-

³ It has been proposed to use instead the term “artificial selection” for artificial systems, originally by John Commons (Commons 1924). However, such a division is mainly semantic: there is no operative distinction between the two and there appears to be no way to tell them apart (Hodgson 2002).

⁴ This effect is often called the “Red Queen Effect” after the Red Queen in *Through the Looking Glass* who tells Alice that to remain in the same place, one has to run and to get anywhere one has to run twice as fast. This captures the arms-race situation quite well,

ner's dilemma model of negative externalities that has the same result as the "tragedy of the commons" game but without a time delay. Agents are faced with a choice of either A or B where A gives the agent a utility of 1 and B gives it a utility of 2 but at the same time all agents in the game also receive a negative utility of $-\varepsilon$. If there is no way of enforcing a collective choice of A, then there is no situation where it is rational to choose A. If B is chosen by all agents then if the number of agents is greater than $1/\varepsilon$ all agents will have a lower utility than had all of them selected A. If even more agents are added, they will still rationally keep choosing B as utility dwindles even lower⁵.

These features have some important implications for how to build models and how to analyze their output. The contingent property of growth and the propensity for self-reinforcing lock-ins means that the details of the future are to a large extent unknowable, they are shrouded in Knightian uncertainty (Knight 1921). Attempts to get around this tends to result in the serving of the same dish on different plates. Either we observe a single future of the system, in which case we get a future which is not necessarily representative of the broader set of possible future states. For instance, say a cluster of development appeared in some place in a cellular automaton model but we know it could just as well have appeared someplace else and once it did show up where it did, the rest of the run was shaped around this "frozen accident". If we on the other hand attempt to average over a Monte Carlo set of runs we get a smooth spatial average instead that says just as little⁶. In other words, this uncertainty is a fundamental property of urban systems and other systems like it⁷. These properties of complex systems models is of course also the reason why orthodox economists dislike them, and it is indeed a good reason. It would, however, have been an even better reason if an alternative had been offered: as it stands the rigorous neoclassical models have exactly the same shortcomings only in yet a different form. They have chosen to simplify away (and thus conceal) this unpredictability already in the axioms of the theory. Thus, the unique outcomes that are achieved say even less about the dynamics of the system. Nelson and Winter refers to such results as "fool's gold" to illustrate that stringency and uniqueness often comes at an unacceptable cost of realism (Nelson and Winter 1982).

The word "urban" enjoys considerably more stability over time than what can be said for the phenomenon that it labels. It was discussed earlier why attempts to explain artificial phenomena in single points in time are futile, except possibly in very limited cases. Thus, it must also be concluded that the persistence of the lite-

firms that do not improve do not retain their fitness, they rapidly lag behind and will find themselves increasingly out competed.

⁵ Rationality appears to be rather safe to assume given such a transparent and simple choice - at least we might suspect that many will also in reality fall for such temptations.

⁶ This is not to deny the usefulness of Monte Carlo runs in general, there might conceivably be non-trivial features that appear very often also in path dependent systems.

⁷ This was also discussed by Roger White at the conference where it was shown that models tend to fare better on the large scale than on the small scale. The problem of "frozen accidents" was also discussed although he prefers to call them bifurcations.

ral labels that we put on such phenomena contribute to perpetuating this common fallacy. That this is so is of course highly unpractical to amend, but it should nevertheless be realized; a theory of urban formation dated to the 19th century might have more or less in common with a contemporary theory carrying the same label; the latter might certainly learn from the former but they are not theories of the exact same thing and thus their genealogic relation is not the same as that between fundamental physical theories of different dates. Herbert Simon explains this in the following way: "... a science of artificial phenomena is always in imminent danger of dissolving and vanishing" (Simon 1969). Put in the Red Queen perspective, theories of the artificial must run as fast as they can to remain in the same place. There is nothing that precludes the possibility the we might in fact have to be satisfied with attaining a theory of today's urban phenomena that is a *worse* model than von Thünen's monocentric model from 1826 was for cities of that era.

Then what are really the direct advantages of complex systems models over equilibrium models? First of all, they bring all the previously mentioned problems of lock-ins and so on into focus: they make them visible and thus puts them on the research agenda and with the evolution of the very concept of urban growth over history, it appears that the formulation flexibility that they afford is if anything becoming more and more important. In contrast, when these effects are concealed already in the axioms of a theory, they will not be treated at all. Even if the predictive power of details is not unequivocally higher⁸, such models might at least say something about what it is that is unpredictable and why. Furthermore, that the details can not be predicted does not imply that no prediction whatsoever is possible. We can for example investigate more general properties of the systems, such as how we can change the general shape of the system, the general mix of land uses, and so on. This can be done quantitatively but also interactively and qualitatively in the shape of "games" where the user can rather freely run and stop the model, change settings, change states and visualize the output⁹. This is indeed an important means of analysis since there exists no better device for analyzing complex systems than the human brain; we surely have not to this day devised any machine more apt at dealing with the complex world in which we live - robotics and artificial intelligence, which represent attempts at doing this, are yet far from being on par with human capacity. Even for problems that can be highly formalized such as chess it is only recently and with great effort that computer programs excel against humans¹⁰. Furthermore, there might be various tricks for better characterizing the evolutionary paths of such models; methods that are more sophisticated than simple averaging. Doing this incorporates acting on several scales concurrently (something that is indeed done already to some extent) both in the construction of models and in the interpretation of the results. For example, in the network model discussed here, city sizes were observed on a larger scale than the model operates. In fact, this is a defining property of complex systems: while it is

⁸ Which it might be.

⁹ This important point was brought up during the conference by Juval Portugali.

¹⁰ Some humans that is. I should add that my own chess playing skills were more than matched by a Commodore 64 two decades ago.

possible in very simple physical systems such as ideal gasses to formulate a macro theory from a micro theory¹¹ new objects appear on new scales and introduce new behavior to more complex systems. A microscopic model thus foregoes the opportunity to use redundancy that appear at such higher levels. The network model discussed here for example does not use such higher-level structure in the urban system it simulates, although it could very well be extended to do this. Things tend to look a lot more complex than they really are if they are viewed on too small a scale. For example, a car trip is easy to describe in terms of a car, persons, fuel, roads and so on but would be completely inexplicable on a molecular scale.

References

- Alonso W (1964) *Location and Land Use*. Harvard University Press, Cambridge, MA
- Andersson C, Frenken K, Hellervik A (2006). A complex networks approach to urban growth. Forthcoming in *Environment and Planning A* **38**: 11: 1941-1964
- Andersson C, Hellervik A, Lindgren K (2005). A spatial network explanation for a hierarchy of urban power laws. *Physica A* **345**: 227
- Andersson C, Hellervik A, Lindgren K, Hagson A, Tornberg J (2003). The urban economy as a scale-free network. *Physical Review E* **68** (036124)
- Andersson CT, Rasmussen S, White (2002) Urban settlement transitions. *Environment and Planning B: Planning and Design* **29**: 841-865
- Arthur WB (1994) *Increasing Returns and Path Dependence in the Economy*. Univ. of Michigan Press, Ann Arbor, MI
- Barabási AL, Albert R (1999). Emergence of scaling in random networks. *Science* **286**: 509-512
- Batty M, Longley P (1994) *Fractal Cities: A Geometry of Form and Function*. Academic Press, San Diego, CA
- Benenson I, Torrens P (2004). *Geosimulation: automata-based modeling of urban phenomena*. John Wiley and Sons, Chichester
- Christaller W (1933). *Central Places in Southern Germany*. Prentice Hall, Englewood Cliffs, NJ
- Clarke KC, Hoppen S, Gaydos L (1997). A self-modifying cellular automaton model of historical urbanization in the san francisco bay area. *Environment and Planning B: Planning and Design* **24**: 247-261
- Commons JR (1924) *Legal Foundations of Capitalism*. MacMillan, New York
- Couclelis H (1985). Cellular worlds: a framework for modeling micro-macro dynamics. *Environment and Planning A* **17**: 585-596.
- Fujita M, Krugman P, Venables T (1999) *The Spatial Economy*. MIT Press, Cambridge, MA
- Gabaix X (1999) Zipf's law for cities: an explanation. *Quarterly Journal of Economics* **114**: 739-767
- Gould SJ (2002) *The structure of evolutionary theory*. Harvard University Press, Cambridge, MA
- Hardin G (1960) The competitive exclusion principle. *Science* **131**: 1292-1297

¹¹ The connection between statistical physics and thermodynamics.

- Henderson JV (1974) The size and types of cities. *American Economic Review* **64**: 640-656
- Hodgson GM (2001) *How Economics Forgot History*. Routledge, London
- Hodgson GM (2002) Darwinism in economics: from analogy to ontology. *Journal of Evolutionary Economics* **12**: 259-281
- Hotelling H (1929) Stability in competition. *Economic Journal*, pp 41-57
- Isard W (1960) *Methods of Regional Analysis*. MIT Press, Cambridge, MA
- Knight FH (1921) *Risk, Uncertainty and profit*. (8th edition 1957) Edition Kelley and Millman, New York
- Lane D (2002) Complexity and local interactions: Towards a theory of industrial districts. In: Quadrio AC, Fortis M (eds) *Complexity and Industrial Clusters*. Springer, Berlin
- Makse HA, de Andrade JS, Batty M, Havlin S, Stanley HE (1998) Modeling urban growth patterns with correlated percolation. *Phys. Rev. E* **58**: 7054-7062.
- Makse HA, Havlin S, Stanley HE (1995) Modelling urban growth patterns. *Nature* **377**: 608-612
- Nelson RR, Winter SG (1982) *An Evolutionary Theory of Economic Change*. Belknap Press of Harvard University Press, Cambridge
- Nietzsche F (1887) *On The Genealogy of Morals*. Boni and Liveright, New York (1913 Translated by H. B Samuel)
- Sanders L, Pumain D, Mathian H (1997) Simpop: a multiagent system for the study of urbanism. *Environment and Planning B* **24**: 287-305
- Saviotti PP (1996) *Technological Evolution, Variety and the Economy*. Edward Elgar Publishing, Cheltenham
- Schweitzer F, Steinbrink J (1998) Estimation of megacity growth. *Applied Geography* **18**: 69-81
- Simon HA (1955) On a class of skew distribution functions. *Biometrika* **42**: 425-440
- Simon HA (1969). *The Sciences of the Artificial*. MIT Press, Cambridge, MA
- Sutton J (2005) Gibrat's legacy. *Journal of Economic Literature* **35**: 40-59
- Tobler W (1979) Cellular geography. In: Gale S, Olsson G (eds) *Philosophy in Geography*. D. Reidel Publishing Company, Dordrecht, pp 379-386
- Torrens PM, O'Sullivan D (2001) Cellular automata and urban simulation: Where do we go from here?. *Environment and Planning B* **28**: 163-168.
- Weber A (1997) *Theory of the location of industry*.
- White R, Engelen G (1993) Fractal urban land use patterns: A cellular automata approach. *Environment and Planning A* **25**: 1175-1199.
- Wilson A (1970) *Entropy in Urban and Regional Modeling*. Pion Ltd, London
- Zipf G (1949) *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA

A Model for Asystematic Mobility in Urban Space

Armando Bazzani, Massimiliano Capriotti, Bruno Giorgini, Giuseppina Melchiorre, Sandro Rambaldi, Graziano Servizi, Giorgio Turchetti

Laboratory of Fisica della Città, University of Bologna, Italy

Abstract

We present an agent-based model to simulate the citizens mobility in a urban space. The request of mobility is determined by the "chronotopic areas": i.e. urban areas where time-dependent activities are installed and attract the citizens according to their social categories. The core of the model is a decision mechanism for the agents based on a daily program, which chooses the transportation means and the roads to reach the scheduled chronotopic areas. The decision mechanism depends on some social characters of the agents, on the information at disposal, on the attraction force towards a chronotopos and on some random choices. The daily program can also be upgraded according to the information given to the agents. The finite volume congestion effects are present in the private transportation and in the finite capacity of the public means whereas the crowding in the chronotopic areas causes the extension of the elapsed time in the areas. We present a simulation on the campus of Milano Bicocca University where we take advantage of some experimental observations on the students mobility.

1 Introduction

The problem of modeling complex urban systems has taken advantage from the continuous increasing of the computer performances and from the new theoretical approaches proposed by the Complex Systems Physics (Schweitzer 1997; Parisi et al. 1998; Ebeling et al. 2001). The attention has been focused at small scale spatial and temporal events where the previous global and static models cannot be applied (Lee 1973, 1994) and the microscopic dynamics is an essential point to be taken into account in the models (Batty 1971; Batty et al. 2002). Urban mobility is a paradigmatic problem from this point of view. Even if the *informatics revolution* based on internet technologies tries to reduce the necessity of mobility, the sociologists have pointed out the mobility is basic instinct of the human nature strictly related with the necessity of social activities and with the idea of freedom. Therefore a sustainable urban mobility is a fundamental task for the life quality in the new metropolis. Where at large scales the mobility is dominated by origin-destination schemes, what complicates the dynamics at mesoscopic and small scales are events in which the stochastic individual movements switch from one decision to another as consequence of interactions with the space geometry and

the other individuals (Martinotti 1999; Nuvolati 2002). This unpredictable mobility has been called asystematic mobility and its role has become more and more relevant in the recent years. A new class of models has been proposed to simulate the urban mobility; these models stress the dynamical non-equilibrium characters and the self-organizing properties of urban mobility and use a holistic approach to the problem (Helbing 2001a). Models are needed not only to reproduce the reality but also to enable users to explore the possible states that could be realistic if different knowledge on the reality is provided to the citizens. In such sense the models should reproduced a virtual reality which has all the structures of the known reality and where virtual experiments (i.e. experiments that are not validated by experimental data) can be performed (Helbing 2001b; Helbing et al. 2001). In this paper we present the *Chronotopic* model for urban mobility *Mobilis in Mobile¹* that has been developed in collaboration among mathematical-physicists, urban planners and sociologists to reproduce the effects of asystematic mobility (Bazzani et al. 2003a). The model is based on citizen microscopic dynamics and the concept of Chronotopos (from the combination of two Greek words: place of time) that indicates a macroscopic area in the urban space where time-schedule activities are present and that determines the citizen request of mobility (Bonfiglioli 1997; Giorgini 1997). The citizens are *attracted* by the activities inside the chronotopic area and move using a probabilistic decision mechanisms. The Mobilis model belongs to the class of models that try to introduce behavioral dynamics (Bierlaire et al. 2003) or artificial intelligence systems to process the information given to the walkers (visual information, memory properties, preferred paths etc.). We recall also the STREET model (Schelhorn et al. 1999), models based on the visual syntax approach (Jiang 1998). At the moment the application of these models to realistic situations is still limited by the difficulties of collecting experimental data.

In section 2 we discuss some details of the Mobilis model and in the section 3 we present an application to a realistic situation: the students mobility in the university campus of Milano Bicocca where some experimental observations have been performed (Boffi et al. 2003).

2 The Mobilis in Mobile Model

Mobilis in Mobile is a dynamical model of the integrated citizen mobility based on a probabilistic microscopic approach which tries to reproduce the individual decisions. By using the Bayes-de Finetti definition of probability that is summarized in the Laplace's sentence "*Probability is nothing but the common sense reduced to calculation*", we introduce the probability as the measure of the individual tendency to make a certain choice (i.e. to choose a certain street at a crossroads or to use a transportation means). The probability takes into account the individual characteristic, the destination, the geometry of the urban space and some aspects related to the subjective perception of the urban space. We introduce three differ-

¹ From the motto of Nautilus: Jules Verne *20000 lieues sous les mers* cap. VIII.

ent representations of an urban space related to different scale. The chronotopic representation of urban space is a time-sensitive map based on the identification of the chronotopoi: i.e. macroscopic areas where time-scheduled activities are present which determine the citizens request of mobility. We remark that the chronotopos is defined in a space-time: for example a certain geographical area is a chronotopos only during the day, whereas other areas become chronotopoi during the night, or two different chronotopoi are present in the same place at different times. The concept of chronotopos is scale-invariant: if we consider a whole city, chronotopoi are the Hospital area, the University campus, the financial centers, ..., whereas if we consider a neighborhood, the chronotopoi are identified with certain buildings (for example a shopping center). The citizens are divided into several classes according to some sociological characters like sex, age, work, instruction degree, etc and each class is attracted by the activity of well-defined chronotopic areas depending on its characters. For example a student feels an attraction of the university chronotopos, the areas where cultural or social activities are present or also shopping centers. Then he organized his daily program in the model by taking into account the constraints of the time-schedule of the different chronotopoi (i.e. probably he has to go to the university in the morning to listen some lessons) and start to move in the urban space. Individuals of the other classes do not feel any necessity to go to the university, but of course someone can reach the university just for curiosity or to visit a museum. The Mobilis model takes into account this possibility by introducing the class city-users (Martinotti 1999) whose mobility is similar to a Brownian motion without a precise destination (asystematic mobility). In a certain sense any individual becomes a city-user when he has completed his program or he moves inside a chronotopic area.

At a lower scale we have the path network representation of an urban space: that is an abstract network map of the geometrical space based on the knowledge of the position of critical nodes (bus stations, passages through obstacles, bridges, squares, ...) and their connectivity properties. We say that the nodes *A* and *B* are connected if it is possible to go from *A* to *B* with a single decision: i.e. two successive stations in a metro line are connected or two crossroads are connected if there is a single street in between (see Fig. 1). By using the network path representation an individual defines ideal paths to reach the desired chronotopic area by using algorithms. A network is also used to represent the public transportation routes where the means follow a deterministic dynamics according to a time schedule.

Finally we have a visual representation of the urban space that associate to each road or place some attributes as comfort, architectural beauty, presence of shops, safety, etc. The pedestrians use such visual information to take local decisions of the direction.

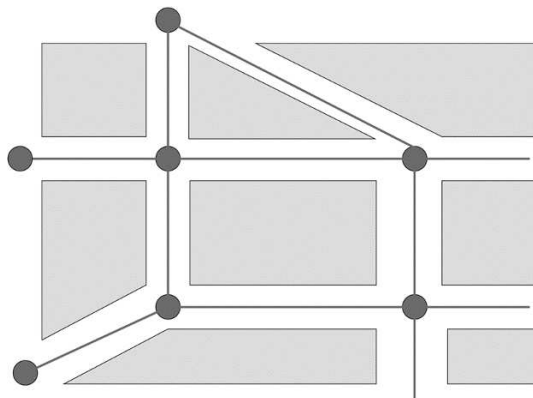


Fig. 1. Path network representation of urban space

Let us consider for example a student that decides to go to the university campus. At the first step he decides if to use a transportation means or to walk towards his destination. At the present state of the model this choice depends on the *Euclidean* distance from the initial point and the university taking into account the individual age and if he is in a hurry. Then he chooses a possible path using the path network representation of the urban space: for example to reach a certain bus station, to take the bus up to another station and then to walk in a certain direction. The path defines sequences of local destinations and the student moves using a simple strategy. The chronotopic attraction defines a vector \mathbf{F} (chronotopic force) whose length is a measure of the tendency to reach the chronotopos and whose direction gives the position the first local destination in the chosen path. We require the constraint $\|\mathbf{F}\| \leq 1$. Arriving at a crossroads we introduce the quantities

$$\pi_j = (1 + \mathbf{F} \cdot \hat{e}_j) \quad (1)$$

for each road j , where \hat{e}_j is a unit vector that gives the direction of the road and \cdot denotes the usual scalar product (see Fig. 2). Then we combine the visual attributes of each road with the sociologic characters of the student: for example the student could not care about the presence of shops in a road or if the pavement is comfortable, but he is more interested in the presence of social or cultural activities. This procedure allows to define a weight w_j for each road and we introduce the probability p_j to choose the road j as

$$p_j = \frac{w_j \pi_j}{\sum_{i \neq k} w_i \pi_i} \quad (2)$$

$$p_k = 0$$

if k is the index of the incoming road. The effective choice is computed by realizing a random variable with probabilities (2b).

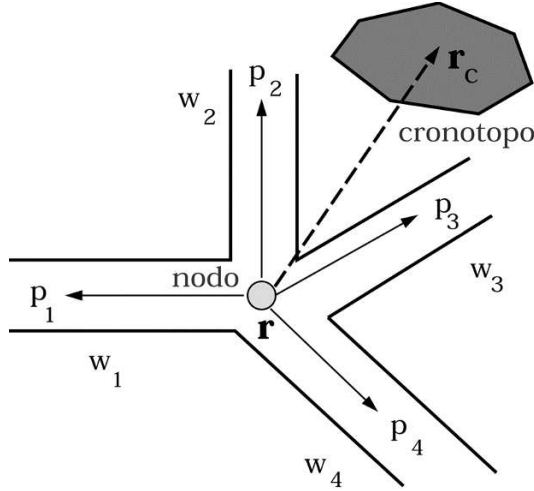


Fig. 2. Schematic picture of the individual dynamics at a crossroads

We observe that when $\|F\| = 1$ the student tends to follow a direct logic path towards his destination (origin-destination mobility), whereas when $\|F\| < 1$ the student has the possibility of illogical random choices and the motion has a Brownian component (asystematic mobility). Along the roads the student walk at a constant velocity but he can also stop with a probability $(1 - \|F\|)/2$ to simulate the possibility of looking windows or of meeting friends if he is not in a hurry. When a local destination is reached the direction of the chronotopic force changes to the next local destination in the path. At the public means stations the individuals stop and expect for the passage of a suitable line. We have introduced a finite capacity for the public means so that the waiting time increases during in a crowded situation. When the chronotopos (university campus) is reached the *force* F is set to zero and the probabilities of Eq. 2 define a random walk, but our student has also to stop for a certain time to perform some activities (for example to listen to the lessons); in such case he disappears from the mobility simulation of the model. The elapsed time τ in a chronotopos depends on the activities in the chronotopos itself and it is sensitive to the crowding effects. We introduce the following definition

$$\tau = \tau_0 + \Delta\tau_0 \frac{N}{N_0} \quad (3)$$

where the quantities τ_0 , $\Delta\tau_0$, N_0 depend on the chronotopos and N is the number of citizens in the chronotopic area. τ_0 gives the time interval required by the chronotopic activities and the ratio $\sigma_0 = \Delta\tau_0/N_0$ defines the chronotopos sensitivity to the crowding effects. The pedestrian dynamics in the model is a stochastic process of the form

$$\begin{aligned} \mathbf{x}(t + \Delta t) &= \mathbf{x}(t) + \hat{e}_v(t)v\Delta t \\ \hat{e}_v(t + \Delta t) &= \xi(\mathbf{x}) \end{aligned} \quad (4)$$

where Δt is the integration step and ξ is a random variable distributed according to the choice probabilities (2b) at the crossroads or constant along the streets. The dynamics is deterministic when the citizens use the public means transportation with the possibility of different choice at the stations. Finally the private cars have more complex dynamics since they obey to the traffic rules and they have relevant physical interactions (Schreckenberg et al. 1995); moreover one should take into account the problem of parking. The time dependence of the population in a chronotopic area is the result of the various citizens class dynamics. In the Fig. 3 we show an example in a Manhattan like city with two chronotopoi and three different classes of agents: in the first (resp. second) class the agents are attracted by the chronotopos A (resp. B), whereas in the third class they have to go first in the chronotopos A and then in the chronotopos B. We observe that both the populations have a smooth initial increase then we have a relaxation to a steady state after a maximum and a sudden decrease when the chronotopos activity is switched off. The chronotopos A reaches its maximum before the chronotopos B since the third class people start to reach later the second chronotopos. The sudden decrease is due to the discontinuity at the border of a chronotopic area.

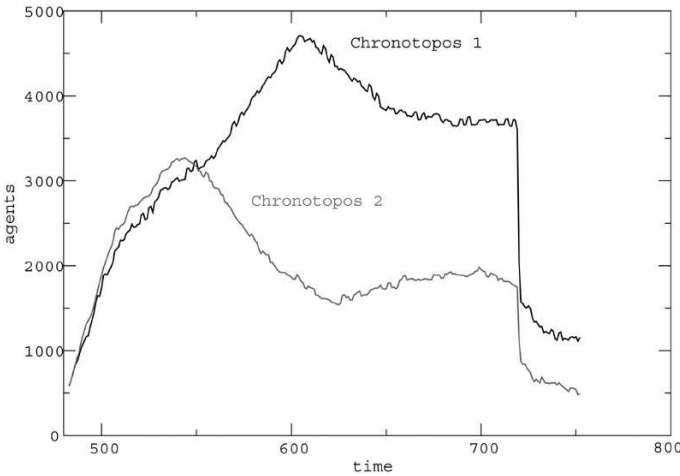


Fig. 3. Example of chronotopos populations in a simulation with 3 different agent class in a Manhattan like city: the first (resp. second) class agents are attracted by the chronotopos A (resp. B), whereas in the third class they have to go first in the chronotopos A and then in the chronotopos B. The time units are arbitrary.

The aim of the Mobilis model is also to introduce an intelligent behavior in the agent dynamics (Von Neumann 1963). We shall discuss in this paper a possible mechanism for the rise of a habit. Let us suppose the agents have to organize their daily programs in order to visit two chronotopoi A and B. Let p_A (resp p_B) the

apriori probability to choose first the chronotopos A (resp. B); due to the effect 3 in the chronotopic area an agent could be not satisfied of his choice. Then we introduce the possibility to change the probabilities according to the Markov scheme

$$\begin{cases} p'_A = p_A(1 - P_{AB}) + p_B P_{BA} \\ p'_B = p_B(1 - P_{BA}) + p_A P_{AB} \end{cases} \quad (5)$$

where P_{AB} (resp P_{BA}) is the transition probability from the choice A to B (resp. B to A). We introduce a dependence of the transition probabilities from a *satisfaction degree* of citizens related to the increase of the elapsed time in the chronotopic areas due to crowding effects. We set

$$P_{AB} = \frac{k}{2}(1 - cn_A + n_B^2) \quad P_{BA} = \frac{k}{2}(1 - cn_B + n_A^2) \quad (6)$$

where $n_A = \sigma_A p_A$ is the product of the probability p_A of choosing the chronotopos A and the sensitivity σ_A to crowding effects (n_B is defined in a similar way). The parameter $k < 1$ measures the citizen attitude to change the previous choice. In the Fig. 4 we plot P_{AB} for $k = .5$ and $c = .65$ to show role of the parameter $c > 0$: when the chronotopos population increases we have an initial decrease of the transition probability p_{AB} (cooperative behavior) then after the threshold $n_A = c$ the transition probability increases with respect to the initial value $k/2$ (non-cooperative behavior).

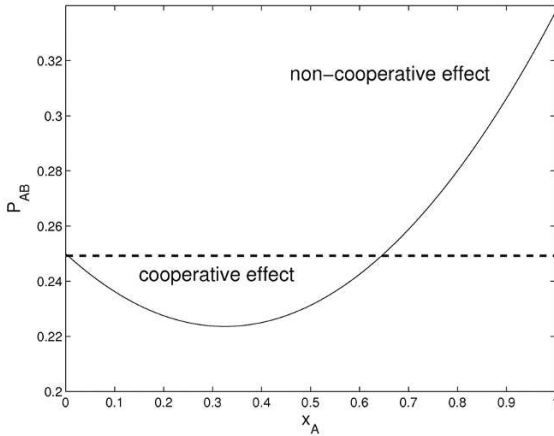


Fig. 4. Plot of the transition probability P_{AB} as a function of n_A for $k = .5$ and $c = .65$; the existence of a cooperative and non-cooperative effect is pointed out.

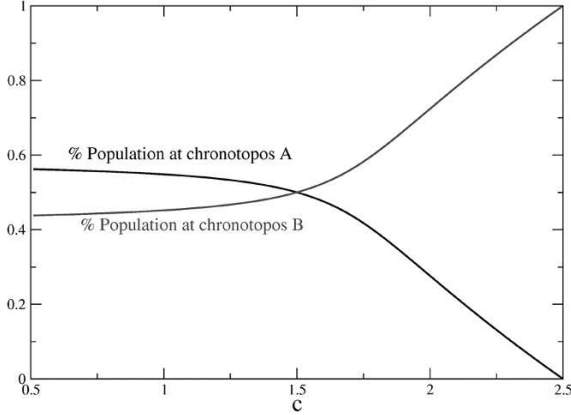


Fig. 5. Equilibrium solutions for the population in two chronotopoi as a function of the cooperation parameter c (see Eq. 7); the chronotopos A has a less sensitivity to crowding effects with respect to the chronotopos B $\sigma_B = 2\sigma_A$ and it is preferred in a non-cooperative situation. On the contrary a strong cooperative behaviour $c > 1$ favours the chronotopos B.

The parameter c measures the citizens attitude to have a social behavior; when $c = 0$ we have completely selfish citizens. Using the Markov scheme (Eq. 5) it is possible to compute the self-consistent equilibrium populations of the chronotopoi A and B as the solution of an algebraic system

$$\begin{cases} p_A = \frac{P_{BA}(\sigma_B p_B)}{P_{AB}(\sigma_A p_A) + P_{BA}(\sigma_B p_B)} \\ p_B = \frac{P_{AB}(\sigma_A p_A)}{P_{AB}(\sigma_A p_A) + P_{BA}(\sigma_B p_B)} \end{cases} \quad (7)$$

In Fig. 4 we show the dependence of the equilibrium solutions p_A and p_B on the cooperation parameter c for 2 chronotopoi with different crowding sensitivity ($\sigma_A = 1$ and $\sigma_B = 2$): we observe that a non-cooperative behavior ($c \ll 1$) favours the chronotopos A, but when cooperation is strong, people prefers to chronotopos B. The behavior could be observed in a case where the chronotopoi are related to social activities. A complete study of the system points out the existence also of a second equilibrium that favours chronotopos A, but its stability basin is much smaller than the basin of the solution plotted in Fig. 5. In a *realistic* mobility simulation one should attribute different decision dynamics to different citizen categories according to sociological studies: for example the housewives could cooperate more than business-men. The introduction of a decision dynamics is also strictly related to mechanisms of information distribution to the population. Indeed any decision dynamics is based on information processes and the role of information is fundamental for the city e-governance. From a theoretical point of

view a selfish behavior and instantaneous complete information on the city dynamical state allow the system to self-organize in an *optimal* way (for example minimizing the elapsed time in the chronotopic areas). This is illustrated in the Figs. 6 and 7 where we plot the population inside 2 chronotopic areas with different properties: the chronotopic activities require $\tau_A = 200$ and $\tau_B = 300$ time units and $\sigma_B = 2\sigma_A$. In Fig. 6 we show the results for a simulation of 16,000 in a Manhattan like city that equally distributed among the chronotopoi A and B: the people who have chosen to visit first the chronotopos A, succeed to perform their duties and then reach the chronotopos B, whereas the people who have first chosen the chronotopos B are in a critical situation since they can't perform their activities before the closing time at 500 time units. Indeed the chronotopos A remains empty after 250 time steps.

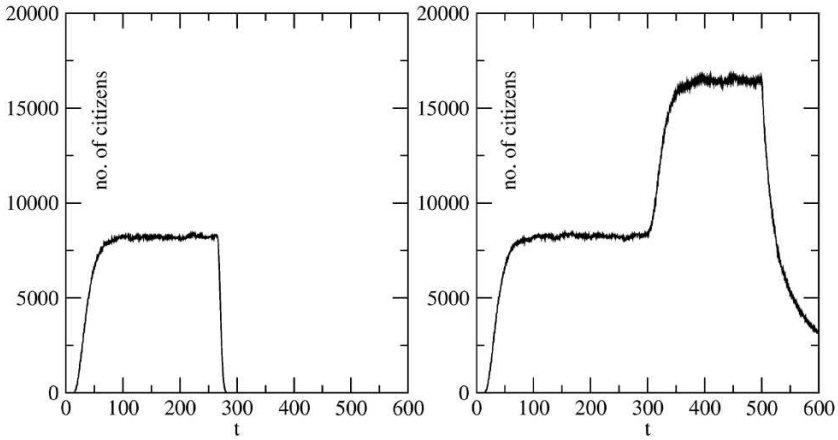


Fig. 6. Chronotopos populations in a Manhattan like city: according to Eq. 3, the activities in the chronotopoi A and B require $\tau_A = 200$ and $\tau_B = 300$ time units whereas the sensitivity to crowding effects satisfies $\sigma_B = 2\sigma_A$ (see Eq. 3). The left picture refers to the chronotopos A whereas the right picture refers to the chronotopos B. The citizens organize randomly their daily programs in order to visit both the chronotopoi. The total number of citizens is 16,000 and at $t = 500$ the chronotopoi activities are switched off. We observe that all the citizens that have chosen to visit first the chronotopos B, do not succeed to complete their activities in the chronotopos A.

By introducing a non-cooperative dynamics that allows to change the daily programs using the information on the chronotopoi populations of the previous day. The system relaxes to a new stationary state (see Fig. 7), where most of people choose to visit first the chronotopos A and then the chronotopos B; in this way the number of people that succeeds to complete the daily program is maximized.

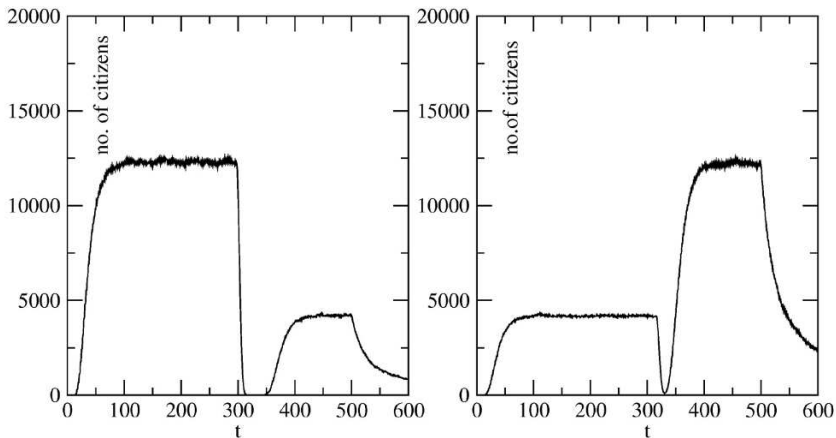


Fig. 7. Chronotopos populations in a Manhattan like city: the parameters used in the simulation are the same as in the Fig. 5. The citizens organize their daily programs in order to visit both the chronotopoi according to a non-cooperative decision dynamics using the information on the populations at the chronotopoi of the previous day. The pictures show the chronotopoi populations after the system reaches a new stationary state: most of the people choose to visit first the chronotopos A so that the number of people that succeeds to complete the daily program is maximized.

3 Applications to Realistic Situations and Experiments

A physical mathematical model of a complex system has the goal to create a *simplified virtual environment* whose dynamical emergent properties and relevant control parameters have a strong analogy with the properties and the parameters of the real system (Bazzani et al. 2006). From this point of view we do not require that a complex system model has to be validated by using Galilean experiments (i.e. reproducible) and gives a faithful reproduction of the reality (Batty 2005). Indeed such a plan will be immediately frustrated by the enormous amount of data necessary for an accurate description of a complex system. In short any *experiment* on a complex system is a unique realization of the system that cannot be reproduced in its details. However any physical model has to be compared to realistic situations in order to understand its capacity to reproduce the interested phenomena and to explore possible scenarios using virtual experiments. Moreover the experimental data can be used to check if the microscopic dynamics that defines the agents' behavior at small scales, are consistent with the observations. The Mobilis model has to face these difficulties in order to be accepted as an instrument to understand the urban mobility dynamics.

The research group coordinated by Guido Martinotti at the sociology department of Milano Bicocca University has built a system to measure the pedestrian mobility in a urban space (Boffi 2003) using the signals of GPS recorders. This system has been used to follow a sample of ≈ 100 students in the Milano Bicocca campus from their entry points (for example the train station, the bus stations, the parking areas or the departments) up to their destinations during three working days of December 2004. The Milano Bicocca campus is an area which is crowded only during the day since there are no residences inside. In Fig. 8 we show a schematic representation of the Milano Bicocca campus with the main chronotopi that determine the mobility request.

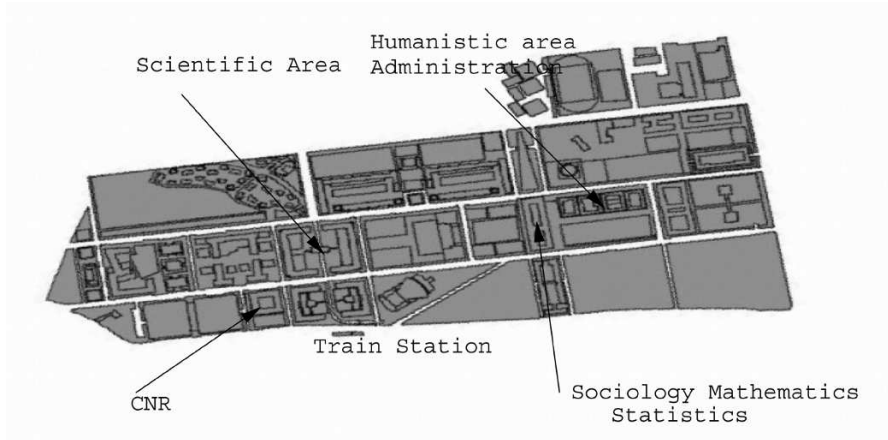


Fig. 8. Schematic plan of the university campus of Milano Bicocca; the main chronotopi are pointed out.

Usually people arrive in the morning and leave the area at lunch time or in the afternoon. The experimental data define the pedestrian paths with a precision of the order of 10 m in the position and a time interval of 30 sec. A histogram of the velocities point out the presence of three types of pedestrian mobility: a slow mobility that could be associated to clusterized students moving together, a mean speed mobility and a high speed mobility probably associated to people in a hurry (see Fig. 9). Therefore we have inserted in the model Mobilis the possibility for the agents to move at different velocities.

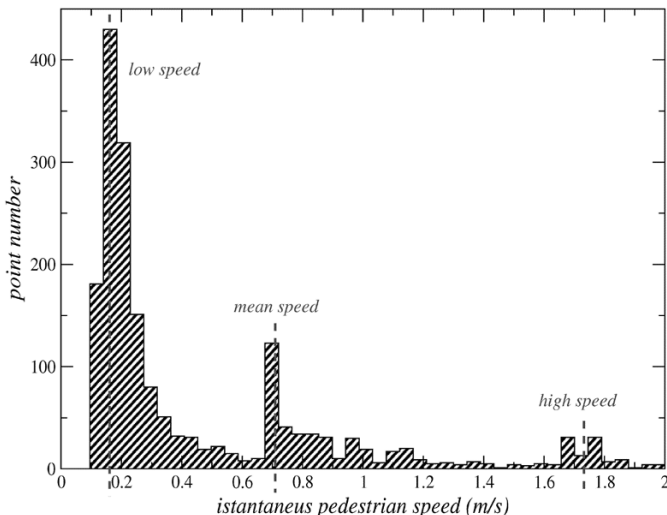


Fig. 9. Histogram of the pedestrian instantaneous velocities measured in the Milano Bicocca campus using all the paths.

In order to perform a virtual experiment we have used the statistical data on the population at the Milano Bicocca campus provided by the sociology department (Bazzani et al. 2003b). We have considered the possibilities of using the train (most popular), buses or private cars to reach the campus according to the statistical data and we have distributed three kinds of daily programs to the population. People who spent only the morning in the Bicocca area, have a program containing a single chronotopos (like a department or a laboratory), whereas people who remain in the area also during the afternoon, have two or three chronotopoi (for example they have to go in a canteen at lunch time and then to return in a department). We have analyzed the experimental data by using an integrated density function to reproduce the distribution of people in the Bicocca area at different times during a working day. In the Fig. 10 we show the integrated density for the experimental paths during the morning between 12.00 and 14.00. We observe that mobility is mainly restricted in the streets between the train station and the scientific and humanistic departments. The peaks correspond to the train station and to the bus stations where people is leaving, but there are also people moving in the campus towards the canteen in the humanistic area. In Fig. 11 we plot the results of a virtual experiment using the Mobilis model for the mobility at noon. The different colours represent the density of moving people along the streets (dark colours indicate a higher density); the active chronotopoi are coloured in black. The virtual experiment seems to well reproduce the mobility observed in the real experiment.

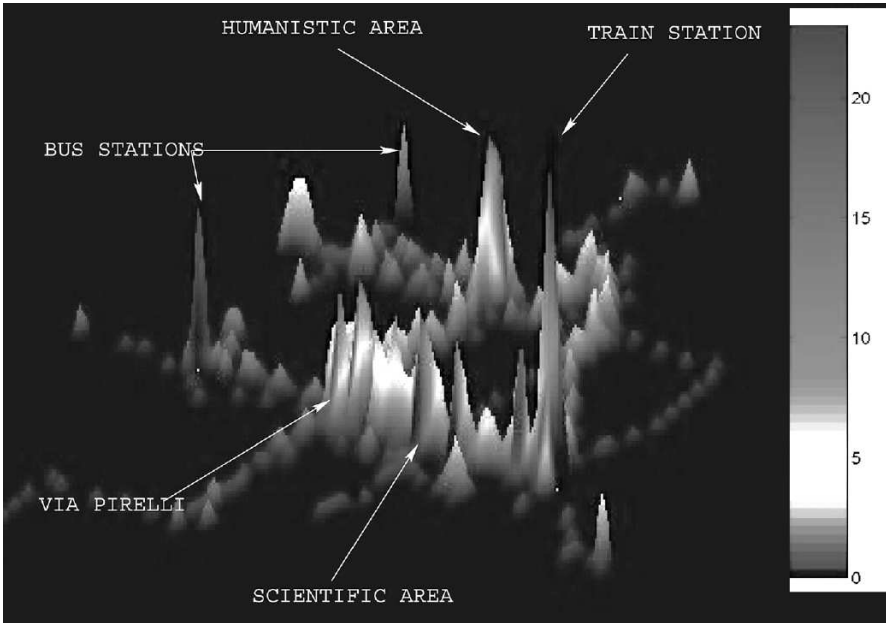


Fig. 10. Integrated pedestrian density from GPS data in the Milano Bicocca campus between 12.00 and 14.00 o'clock..

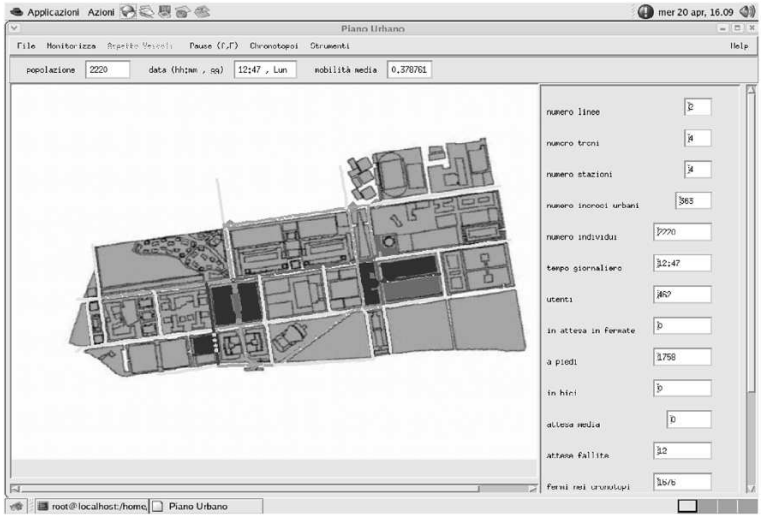


Fig. 11. Simulation of the mobility at noon in the Milano-Bicocca campus using the Mobilis model. The different colours represent the density of moving people along streets (dark colours indicate a higher density); the active chronotopoi are coloured in black.

4 Conclusions

The agent based models like Mobilis in Mobile seem to be suitable to reproduce the new kinds of mobility that have been observed in the new generation metropolis. From one side these models simulate the small scale unpredictable events due to the individual freedom and from the other consider the urban mobility from a holistic point of view typical of complex systems. The comparison of the simulations with realistic cases requires new approaches to the validation problem due to the difficulty of performing experiments on complex systems. The future goal is to create a virtual reality that explores all the possible scenarios in the considered case.

Acknowledgments

We are in debt with Mario Boffi and Guido Martinotti for very helpful discussions on the Milano Bicocca experiments. This work was partially supported by Volvo Italia.

References

- Batty M (1971) Modeling cities as dynamics systems. *Nature* **231**: 425-428
- Batty M, DeSyllas J, Duxbury E (2002) The discrete dynamics of small scales spatial events: agent based models of mobility in carnivals and street parades.. http://www.casa.ac.uk/working_papers/paper56.pdf
- Batty M (2003) Agent-Based Pedestrian Modeling. In Longley PA, Batty M (eds) *Advanced Spatial Analysis*. ESRI Press, Redlands, CA
- Batty M (2005) *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. MIT press, Cambridge, MA
- Bazzani A, Giorgini B, Servizi G, Turchetti G (2003a) A Chronotopic Model of Mobility in Urban Spaces. *Physica A* **325**: 517-530
- Bazzani A, Capriotti M, Giorgini B, Melchiorre G, Servizi G, Turchetti G, Luccardi S, Zanolini M (2003b) Un modello per la mobilita' asistemica nel Centro Storico di Rimini. *Trasporti & Territorio* **16**: 141-149
- Bazzani A, Giorgini B (2006) The physics of the City: Modeling Complex Mobility. In *Urban civilization: from yesterday to the next day*. Proceedings of the workshop Urban Science Forward Look, Milano 24 September 2004, to be published
- Bierlaire M, Antonini G, Weber M (2003) Behavioral dynamics for pedestrians. 10th International conference on Travel Behavior reserach, Lucerne 10-15 August 2003
- Boffi M, d'Ovidio M, Natoli E, Tornaghi C (2003) Slow Mobility as and Indicator of Social Urban Spaces - An Application on GPS in the Milano-Bicocca Area. In Gould M et al. (eds.) *AGILE 2003*. Proceedings of the 6th Conference on Geographic Infomation Science, PPUR, Lausanne

- Bonfiglioli S (1997) Che cos'è un cronotopo. In: Bonfiglioli S, Mareggi M (eds) *Il tempo e la città fra natura e storia. Atlante di progetti sui tempi della città*, Urbanistica Quaderni INU, no. 12, Roma, p 90
- Ebeling W, Schweitzer F (2001) *Active Motion in Systems with Energy Supply in Integrative Systems Approaches in Natural and Social Sciences*. Springer, Berlin
- Giorgini B (1997) Il tempo del caos. In: Bonfiglioli S, Mareggi M (eds) *Il tempo e la città fra natura e storia. Atlante di progetti sui tempi della città*. Urbanistica Quaderni INU, no. 12, Roma, p 325
- Helbing D (2001a) Traffic and related self-driven many-particle systems. *Review of Modern Physics* **73**: 1067-1141
- Helbing D (2001b) A mathematical model for the behavior of pedestrians. *Behavioral Science* **36**: 298-310
- Helbing D, Molnar P, Farkas I, Bolay K (2001) Self-Organizing pedestrian movement. *Environment and Planning B* **28**: 361-383
- Hoogendoorn SP, Bovy PHL, Daamen W. (2001) Microscopic pedestrian wayfinding and dynamics modeling. In: Schreckenberg M, Sharma SD (eds) *Pedestrian and Evacuation Dynamics*. Springer, Berlin, pp 123-154
- Jiang B (1998) SimPed: Simulating Pedestrian Flows in a Virtual Urban Environment. *Journal of Geographic Information and Decision Analysis* **3**: 1: 21-30
- Lee DB (1973) Requiem for large-scale models. *Journal of the American Institute of Planners* **39**: 163-178
- Lee DB (1994) Retrospective on large-scale urban models. *Journal of the American Institute of Planning Association* **60**: 1: 35-40
- Martinotti G (ed) (1999) *La dimensione metropolitana: sviluppo e governo della nuova città*. Il Mulino, Bologna
- Nuvolati G (2002) *Le popolazioni in movimento e la città in trasformazione*. Il Mulino, Bologna
- Parisi G, Muller SC, Zimmermann W (1998) *A Perspective Look at Nonlinear Media - From Physics to Biology and Social Sciences*. Springer, Berlin
- Schreckenberg M, Schadschneider A, Nagel K, Ito N (1995) Discrete stochastic models for traffic flow. *Physical Review E* **51**:4: 2939-2949
- Schelhorn T, O'Sullivan D, Haklay M, Thurstain-Goodwin M (1999) STREET: an agent-based pedestrian model. Centre for Advanced Spatial Analysis, Paper 9
- Schweitzer F (1997) *Self-Organization of Complex Structures: From Individual to Collective Dynamics*. Gordon and Breach, London, pp 19-24
- Servizi G, Bazzani A, Giorgini B, Turchetti G (2001) Method for simulating mobility in an urban area. Patent n. 00830868.6-2215 (13-9-01)
- Von Neumann J (1963) The general and logical theory of automata. *Collected works vol. V*, Pergamon Press, Oxford, pp 288-328

Preliminary Results of a Multi-Agent Traffic Simulation for Berlin

Ulrike Beuck¹, Marcel Rieser¹, David Strippgen¹, Michael Balmer², Kai Nagel¹

¹ Transport System Planning and Transport Telematics (VSP), TU Berlin, Germany

² Institute for Transport Planning and Systems (IVT), ETH Zurich, Switzerland

Abstract

This paper provides an introduction to multi-agent traffic simulation. Metropolitan regions can consist of several million inhabitants, implying the simulation of several million travelers, which represents a considerable computational challenge. We report on our recent case study of a real-world Berlin scenario. The paper explains computational techniques necessary to achieve results. It turns out that the difficulties there, because of data availability and because of the special situation of Berlin after the reunification, are considerably larger than in previous scenarios that we have treated.

1 Introduction

In recent years, microscopic traffic simulations have become an increasingly active field of research in transport science. “Micro” refers to the fact that all elements of the transport system, like roads, crossings, vehicles, and – most importantly – travelers (referred to as “agents”) are resolved. This modeling approach is in contrast with the more aggregate models implemented in current transport planning software and used by transportation planners. While those programs have seen several decades of development and practical use, agent-based microscopic simulation systems are still relatively new and are mostly used in small and medium scale scientific scenarios rather than in real world applications. But new technologies, such as robust and fast object-oriented programming languages and high performance computing clusters make the applications increasingly realistic and increasingly large scale. This paper addresses the issue of applying such a model to a real world scenario of large dimensions.

2 The Model

2.1 Physical vs. Mental Level

There are several ways from which a microscopic approach can be derived. One way is the attempt, often used in physics, to start from “first principles”. First principles implies to start from individual particles, and indeed the possibility to do fast molecular dynamics (Beazley et al. 1995) and fast cellular automata simulations (Stauffer 1991) on the microscopic level was one of the driving forces for large scale microscopic traffic simulations (Nagel and Rickert 2001).

When going down this path, one notices eventually two things:

- Building a microscopic simulation of vehicular traffic (or, for that matter, of pedestrians) needs some diligence and care, but is essentially possible.
In terms of size: Even large urban systems rarely have more than 10^7 inhabitants, and rarely more than about 20% of those are simultaneously on the road. This makes for considerably smaller numbers of particles than in many physics applications, and moves the microscopic simulation of complete cities/regions into the realm of the computationally feasible.
In terms of underlying dynamics: Despite considerable discussion about different traffic states and possible phase transitions (Kerner et al. 2002; Helbing et al. 1999; Jost and Nagel 2003), the absolutely most important elements of the dynamics are, in fact, rather constrained: There is mass conservation (especially if vehicles are tracked from parking to parking; see below); vehicles move on a quasi one-dimensional geometry (roads) most of the time; there are quite severe restrictions on acceleration and braking capabilities; and there are quite severe excluded volume restrictions (not more than about 150 vehicles fit on a kilometer of single-lane roadway when traffic is stopped). A consequence of this is that a rather simple theory of traffic – that of kinematic waves (Lighthill and Whitham 1955) – describes traffic rather well, by just using the equation of continuity plus an equation of state (relating velocity to density, the so-called “fundamental diagram”) (Nagel and Nelson 2005). Any microscopic model that obeys the corresponding microscopic principles – mass conservation and velocity related to the distance to the car ahead – will reach a similar level of realism (Brockfeld et al. 2003). More complicated aspects, such as capacity drop (Hall and Agyemang-Duah 1991), phase transitions (see above), or “synchronized traffic” (Kerner 1998), matter for the management of individual road segments, but they do not matter so much for where we currently are with simulations of large scale urban systems.
- The other thing that one notices is that the behavior of the “particles” (vehicles, pedestrians) is quite heavily influenced by behavioral aspects, i.e. by “what goes on in people’s heads”. This is, however, not so much the realm of physics, since one is not interested in, say, how 10^6 neurons together eventually make a decision, but instead in models that generate realistic human decisions within very short computing time. For typical computing situations, a lane changing

decision may not take more than 10 s of CPU time, and a decision about a daily plan may not take more than a second of CPU time. That is, one is interested in models that describe the outcome of human decision-making reasonably well, without “looking at the neurons”.

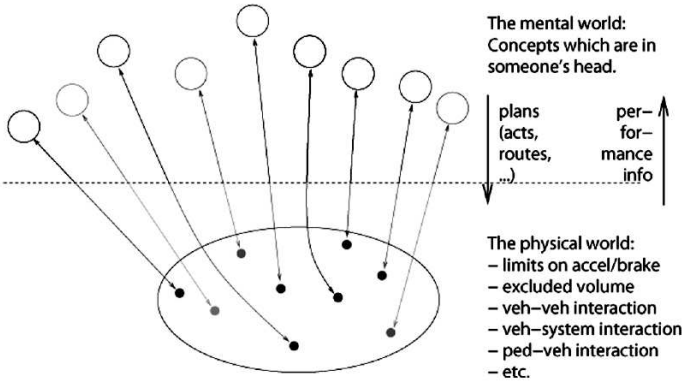


Fig. 1. Physical vs. mental level

Accordingly, it is useful to differentiate between the **physical layer** and the **mental layer** of a multi-agent simulation (MASim), see Fig. 1. The physical layer essentially contains everything that can be observed. The mental layer essentially contains everything that goes on in people's heads.

2.2 The Physical Level

As just said, there is a variety of techniques available for the simulation of the physical level. These techniques include molecular dynamics (Bando et al. 1995), techniques based on differential equations (e.g. Helbing 2001), coupled maps (Gipps 1981; Krauß et al. 1996; Gloor et al. 2003), cellular automata (Chowdhury et al. 2000), methods where individual vehicles are moved with velocities based on link densities (Ben-Akiva et al. 1998; Chang et al. 1985), methods where individual vehicles are moved based on fluid-dynamical equations (Flötteröd and Nagel 2005), and queue models (Gawron 1998). Some packages based on these different techniques are SUMO (using coupled maps and more recently a queue model; SUMO [www page](#)), DYNASMART, DYNAMIT, METROPOLIS (all using a combination of velocities based on link densities and a queuing approach; DYNASMART [www page](#); DYNAMIT [www page](#); de Palma and Marchal 2002), TRANSIMS, OLSIM (both using a cellular automata approach; TRANSIMS [www page](#); Esser 1998), or our own package MATSIM (using a queuing approach and more recently also vehicle movement based on fluid-dynamics; MATSIM [www page](#)). Since these are not the main focus of this paper, it shall suffice to have given these references.

2.3 The Mental Layer

As mentioned above, the mental layer models and simulates the human decision processes. Those include:

- accelerating, braking, lane changing
- turning decisions at intersections; route choice
- time choices (when to depart?)
- mode choices (which mode of transport to use?)
- location choices (where to do an activity?)
- activity pattern choice (which activities should be done at a given day, and in what sequence?)

This is an approximately hierarchical list, in the sense that decisions further down the list are made less often, and in consequence decisions further up the list depend on those further down the list. For example, in order to compute a route from home to work, one first needs to know where home and work are located.

Although the above hierarchy can be justified by empirical observation (e.g. Miller and Roorda 2003), there is also considerable inverse causality between the levels. For example, location choice (for example for a shopping location) depends on the available modes of transport.

With respect to model implementation, the following seems to establish itself in the community:

- Driving behavior, such as accelerating, braking, or lane changing, is included into the physical layer. That is, it is not assumed to be part of any strategy, but rather assumed to be purely reactive.
- Routes are typically generated using some kind of shortest (or fastest) path algorithm. This is most probably due to the fact that a shortest/fastest path in a traffic graph is relatively cheap to compute by using the Dijkstra algorithm (Dijkstra 1959); it is, in fact, difficult to devise heuristics that are faster than that exact algorithm. If several route alternatives are available, selection between them is often done using a so-called multinomial logit or probit model (Ben-Akiva and Lerman 1985). Some care needs to be taken to correctly deal with correlations between alternatives (Caschetta and Papola 1998): Assume three route alternatives, where two of them differ in just one link, and the third is very different. The intuitive split between those would be roughly 25% : 25% : 50%, while plain multinomial logit returns 33.3% : 33.3% : 33.3%. This is known as “independence from irrelevant alternatives (IIA-property)” (see, e.g., Ben-Akiva and Lerman 1985).
- The choices of times, modes, locations, and activity patterns are often done in one model, called activity-based demand generation (ABDG). There are two major strains of models: those based on econometrics/utility maximization, and those based on rules. Most real-world implementations are a combination of those two approaches (e.g. Pendyala, 2005; Bhat et al. 2004; Bowman et al. 1999; Jonnalagadda et al. 2001; Arentze and Timmermans 2000, 2005).

The above assumes that the population and where it lives is given. That is, the (synthetic) population can be seen as a fixed boundary condition of such simulations. Synthetic populations are generated from demographic data (e.g. Beckman et al. 1996). Some models, normally separate from traffic models, also consider the evolution of a population over time (including aging, birth, and death), and include residential choice into those models (e.g. Salvini and Miller 2005; Waddell et al. 2003; Strauch et al. 2002).

In our work, we call the output of the mental layer plans, which can look like:

```
<person id=="123" gender="male" income="50k">
  <plan score="456">
    <act type="home" link="234" endTime="08:05"/>
    <leg mode="car" expectedTravelTime="00:55">
      <route> 25 35 46 63 </route>
    </leg>
    <act type="work" link="345" duration="09:00"/>
    ...
  </plan>
</person>
```

That is, a plan is a full description of the agents' intentions. The above agent intends to leave home at 8:05, take a pre-specified route to work with an expected travel time of 55 min, work for 9 hours, etc. Although in much of our work, plans are fully specified, conceptually they do not need to be so: It is quite reasonable to assume that some elements (e.g. the time to leave work) is decided depending on circumstances (e.g. how much work there is), and other elements are modified on short notice (e.g. the route, in order to circumvent some exceptional congestion).

2.4 Adaptation and Learning; Day-To-Day vs. Within-Day Replanning

If one runs the sequential process of synthetic population generation, activity-based demand generation, and routing, the resulting plans are often not useful since they will not execute as expected. A typical obstacle is congestion, which is a consequence of too many plans attempting to use a certain element of the infrastructure at the same time. Congestion will make certain choices sub-optimal, in the sense that an agent could find a better solution by modifying its plan.

There are two principal ways to model replanning:

- Option 1, called **day-to-day replanning**. The physical level simulates a day, then the plans of the agents are adapted, the physical level simulations a new day based on the new plans, etc. In pseudo-code:

```
for ( day = 1 ; day <= lastDay ; day ++ ) {
  for ( time = 0 ; time < endOfDay ; time ++ ) {
    advance_physical_layer_by_one_time_step ( ) ;
  }
  for ( agent in agents ) {
    agent.replan ( ) ;
  }
}
```

Most of the syntax is hopefully clear; `for (agent in agents)` means that the loop goes through all agents; `agent.replan()` means that the specific agent is now asked to potentially replan.

In this version, agents pre-plan their complete day before they leave home in the morning, and they can only re-consider their plan just before they start the next morning. Agents can therefore not react to unforeseen circumstances.

- Option 2, called **within-day replanning**. The physical level simulates a short time period, then all agents can re-plan, the physical level simulates another short time period, etc. In pseudo-code:

```
for ( day = 1 ; day <= lastDay ; day ++ ) {
    for ( time = 0 ; time < endOfDay ; time ++ ) {
        advance_physical_layer_by_a_one_time_step ( ) ;
        for ( agent in agents ) {
            agent.replan() ;
        }
    }
}
```

More intelligent/efficient implementations of this can be considered, such as agent replanning being triggered by certain conditions during the update of the physical layer.

In this version, agents can replan while they are on the way, thus being able to react to unforeseen circumstances. Important examples of unforeseen circumstances are fluctuations from one day to the next, or exceptional events, such as accidents.

The two options face different levels of implementation difficulties:

- Option 2 is easier to implement by a single programmer or by a tightly integrated programming team where all members of the team have agreed to use the same data structures (e.g. for the agents).
- Option 1 is easier to implement if there is pre-existing, non-integrated code, or if the programming team is not tightly integrated.

In consequence, within-day replanning is often implemented by single-person projects (Emmerink 1996; de Palma and Marchal 2002; SUMO [www page](#); Flötteröd and Nagel 2006) or by projects that can define and enforce their programming standards, while day-to-day replanning is often the result of a multi-person or multi-team project (DYNAMIT [www page](#); TRANSIMS [www page](#); Strauch et al. 2002).

A direct consequence is that projects with within-day replanning are often somewhat limited in scope, since it is difficult to combine pre-existing work. Since, on the other hand, within-day replanning is an important aspect of reality, it seems critical to overcome that obstacle.

2.5 Scores and “Events”

For most applications, replanning only makes sense if the agent attempts to obtain a “better” plan by replanning. This implies that one needs to be able to compare plans. We assume that the plan is scored by submitting it to the physical layer, and scoring the outcome. That is, the plan is seen as the description of the strategy of the agent. The strategy is then interpreted and executed by the physical layer. If the strategy contains infeasible elements (e.g. a route that is not connected), it will fail completely. Even if the strategy is feasible, it may not be very good, since, for example, assumptions about travel times may be optimistic. The physical layer provides output in terms of events, which structurally look as follows:

```
<event time="08:05" agentID="123" eventType="leavingAct"
act = "home"/>
<event time="10:00" agentID="123" eventType="arrivingAtAct"
act = "work"/>
<event time="19:00" agentID="123" eventType="leavingAct"
act = "work"/>
```

This refers to the agent as described earlier by the example plan. The agent leaves home as intended, but needs one more hour than intended to get to work. Consequentially, she will leave work one hour later than intended, since the duration of work is fixed by the plan. The scoring will be based on the longer duration of travel, and the later work start/work end times. If, as a result, she gets late to an appointment later, that will cause further score reduction. If the morning travel delay occurs regularly, she will learn, say, to depart earlier.

Scoring functions can be arbitrary, although in practice, it is currently easiest to remain close to the utility functions used in economics. There is some research into how to assign utilities to full daily plans (Jara-Diaz et al. 2003).

2.6 Co-Evolution, Dynamical Systems, and Evolutionary Game Theory

Day-to-day learning can be described in terms of an evolutionary game. If, over the iterations, all agents end up with plans that they cannot unilaterally improve, then the system is at a **Nash equilibrium**. The plans can be seen as “strategies”; the execution of the plans in the physical level can be seen as “scoring the strategies” or “computing the fitness function”. The concept even holds when the simulation of the physical level is stochastic; then “score” needs to be replaced by “expected score”.

The day-to-day evolution of the system can be seen as a time-discrete dynamical system where many agents co-evolve. For such systems, some theory is available (Hofbauer and Sigmund 1998), although there is a gap between theoretically tractable systems in the traffic context and full-blown multi-agent simulations. For such theoretically tractable systems, one can show: If, in every iteration (day), a small fraction of the agents switches to what would have been the best expected

plan in previous iterations (“**best reply**”), and if that system moves to a fixed point, than that fixed point is a Nash Equilibrium. There are, however, at least two caveats: (1) The implementation of “best expected score” is not easy to make operational because a simulation is computationally rather expensive, and averaging over several such simulations is even worse. (2) As is well known, dynamical systems do not need to converge to fixed points. They can instead converge to cycles, or to chaotic attractors (Schuster 1995).

3 Set-Up of the Berlin Simulations

Although the scenario focuses on the city of Berlin, in order to simulate average traffic conditions we have to model and simulate Berlin’s surrounding as well but with a lower level of detail. All together the study region covers an area of 150 km x 250 km and has a population of about 6 million inhabitants. Network and demand are derived from data used and produced by the aggregated macroscopic model that Berlin’s planning department is working with. In contrast to this official transport model used for mid and long term forecasts, in our simulations all travelers are resolved as agents generating trips while following their day plans. The following sub-sections describe the set-up of the Berlin scenario.

3.1 Boundary Conditions: The Network

As mentioned earlier, Berlin’s planning department provided us with a road network of their transport model. This network has been used as part of the forecast model for the year 2015. Since we aim to model and simulate Berlin’s current traffic of an average workday, we had to adapt the network manually in order to exclude modifications planned to be realized until 2015 (e.g. expansion of the inner city highway southward). The final network consists of almost 30,000 links connecting more than 10,000 nodes, described by their coordinates. For our simulation we need, for each link, the attributes free flow speed, length, number of lanes and flow capacity. The network does indeed contain these attributes, but the usefulness of the data is variable. For example, the number of lanes is uniformly set to one, presumably because the number of lanes does not matter for traditional assignment models. Link capacity is interpreted very differently by the aggregated model used by the planning department of Berlin and our multi-agent simulation. While in our simulation, capacity is understood as maximum outflow of a link in a given time period, the aggregated model does not treat a link’s capacity as hard constraint. In traffic assignment suitable functions are used to relate capacity and flow with the resulting cost in terms of travel times. Thus, we had to adapt these capacity values that were the basis for a 24 hours static assignment. In a first step, we adjusted the 24 hours capacity values in order to derive 1-hour values based on the assumption that daily traffic basically occurs in a 12-hours period. In a second

step, we converted the resulting theoretical 1-hour values into maximum values of outflow of a link in 1 hour to be used in our multi-agent simulation.

Free flow travel time is calculated as link length divided by the free flow speed of the link. Additionally, the storage of a link is constrained. The storage of a link is calculated as length times the number of lanes divided by the space a vehicle occupies in a jam (7.5 m). Because of the incorrect number of lanes (uniformly one, which is much too small for the wide roads of Berlin), the space capacity needed correction as well. For the time being, we assume a storage of 3-lane roads everywhere – note that this affects only the storage (maximum number of vehicles on link), not the flow capacity.

In order to speed up the Berlin scenario, the demand and the network capacities (both flow and storage) were scaled down to 10% of the actual values.

3.2 Initial Plans from an Activity-Based Demand Generation Program

Initial plans have their source in an activity-based demand generation (ABDG) model (also known as Berliner Personenverkehrs-Modell; Kutter and Mikota 1990; Kutter 1984; Kutter et al. 2002). It has been used to calculate three daily (= 24-hour) OD-matrices used as input data for the static assignment used by Berlin's planning department, differentiated between personal travel, freight travel, and through traffic. However, the model is in fact a disaggregated activity- and behavior-oriented traffic demand generation model. The demand of 72 person groups with similar demographic attributes and homogeneous behaviors is calculated based on expectancies. The model was modified to output activity chains to be used to produce initial agents' plans for our multi-agent simulation (Rümenapp and Steinmeyer 2006).

Activity chains are grouped by the 72 person groups. Each activity chain contains information about the start location, up to four activities, and the frequency of occurrence of the activity chain. Activities are described by their type, location, and the transportation mode used to reach that location. The home location is start and end location of each activity chain (round trips). Information on location refers to traffic analysis zones (TAZ), since these represent sources and sinks of traffic streams in the macroscopic model. Before transforming activity chains into agents' plans, location information and data has to be disaggregated. Additionally, activity chains lack time information. For initial plans, all activities are assigned a random activity duration within a type-dependent range. As a result, over 7 million "virtual" agents are generated from the round-trips in the ABDG data. Each of these agents has a plan corresponding to an activity chain generated by the Kutter model. Unfortunately, the number of these agents does not correspond to the number of real persons in the area, since persons who make more than one round trip per day are registered as separate "virtual" agents. This is due to the fact that the Kutter model treats round trips, not persons. That is, activity chains with intermediate home stops are treated as completely separate round trips, resulting in separate agents.

We then decreased the number of agents in our simulation to the agents using the car for transportation. As already mentioned, to speed up the Berlin scenario we also scaled network capacities as well as demand down to 10, which gives a total of about 205,000 car travelers with complete day plans in our simulation.

3.3 Mental Layer: “Planomat” and Router

As already described, an agent’s plan is a description of its intentions, but it might not be executed as expected because of congestion effects. At the end of an iteration, a score is calculated for each plan, corresponding to how successful an agent was performing its plan (see Sec.3.5). A certain percentage of agents can adapt their plans, before the next simulation of the physical layer starts.

Two strategy modules enable this day-to-day replanning. The first module is the router. Given locations, departure times and activity durations, an agent tries to find a better route in terms of minimum travel costs based on the previous iteration. The router is based on Dijkstra’s shortest-path algorithm, and shortness is measured by travel costs in terms of travel times on the links of a route. Travel times depend on how congested the links are, and so they change throughout the day. The second strategy module is the so-called planomat (Meister et al. 2006). Using this module, departure times or activity durations can be altered in order to optimize the score of the plan. Also, altering the activity sequencing and activity dropping are possible modifications but are not implemented yet.

3.4 Physical Layer: Queue Simulation

The physical layer is simulated using a queuing approach (Gawron 1998). The agents’ plans are executed, and according to the plans they are moved on the network. As output of the simulation, events are produced allowing to calculate travel time, speed, etc. In general, an agent is moved to the next link when it was on that link for at least the free flow travel time, according to the maximum outflow, and when there is space on the next lane. The mentioned networks attributes remain fixed, as mentioned above. More information about the queue simulation can be found in (Cetin et al. 2003).

3.5 Scoring

Scoring a plan is a precondition so that agents learn. Different plans can be compared and an agent can pick the one with the highest value. A higher score implies that the agent makes better use of its day. A scoring function needs to be defined, which evaluates complete day plans. As scoring function, the traditional utility function based on the Vickrey bottleneck model is used (Arnott et al. 1993), but modified to be consistent with complete day plans. Scoring is based on events in-

formation from the physical layer. Performing an activity is rewarded, travel times and late arrival are punished. The overall equation is:

$$U_{plan} = \sum_i U_{act,i} + \sum_i U_{trav,i} + \sum_i U_{late,i} \quad (1)$$

We assume the utility of performing an activity as increasing logarithmically:

$$U_{act,i}(x) = \max \left[0, \alpha \cdot \ln \left(\frac{x}{t_0} \right) \right] \quad (2)$$

where x is the duration that one spends at the activity. We take $\alpha = \beta_{dur} t^*$ where β_{dur} is uniformly the same for all activities and only t^* varies between activity types. With this formulation, t^* can be interpreted as a “typical” duration, and β_{dur} as the marginal utility at that typical duration:

$$\left. \frac{\partial U_{act,i}}{\partial x} \right|_{x=t^*} = \beta_{dur} \cdot t^* \cdot \frac{1}{t^*} = \beta_{dur} \quad (3)$$

t_0 can be seen as a minimum duration of an activity, but is better interpreted as a priority: All other things being equal, activities with large t_0 are less likely to be dropped than activities with small t_0 . For details, see Charypar and Nagel (2005).

The utilities of traveling and of being late are both seen as disutilities which are linear in time:

$$U_{trav,i}(x) = \beta_{trav} \cdot x \quad (4)$$

(where x is the time spent traveling) and

$$U_{late,i}(x) = \beta_{late} \cdot x \quad (5)$$

(where x is the time an agent arrives late at an activity). β_{trav} is set to -6 €/h , and β_{late} is set to -18 €/h .

In principle, arriving early or leaving early could also be punished. There is, however, no immediate need to punish early arrival, since waiting times are already indirectly punished by foregoing the reward that could be accumulated by doing an activity instead (opportunity cost). In consequence, the effective (dis)utility of waiting is already -6 €/h .

Similarly, that opportunity cost has to be added to the time spent traveling, arriving at an effective (dis)utility of traveling of -12 €/h .

No opportunity cost needs to be added to late arrivals, because the late arrival time is already spent somewhere else. In consequence, the effective (dis)utility of arriving late remains at -18 €/h .

These effective values are the standard values of the Vickrey model (Arnott et al. 1993).

It would make sense to consider an additional punishment (negative reward) for leaving an activity early. This would describe, for example, the effect when there are, on a specific day, better things to do than to continue to work, but some kind of contract (e.g. shop opening hours) forces the agent to remain at work.

If a new plan is built, an agent will execute it in the next iteration in order to obtain a score. In general every plan is scored after being executed. A formerly good plan can be scored at lower values if conditions change, e.g. congestion effects.

3.6 Details of the Learning Algorithm

The simulation starts with initial plans. Executing all agents' plans simultaneously gives the agents' interactions in the network. By allowing the agents to re-adjust, they can learn from the previous iteration (feedback learning). The iterations will go on until the system does not show any further development. In other words, agents adapt to their environment and learn how to improve their plans over many iterations. In the simulation all agents learn at the same time, since their plans are executed simultaneously. This also means, that an agent's environment changes due to the effect of the other agents in the system. Thus a plan's score has to be updated.

An agent database keeps track of agents and their decisions, allowing them to choose a strategy based on their past actions. An agent can compare plans of its repertoire by the score they got in previous iterations. In the course of the simulation, the agents learn to build good plans in order to realize their intentions and to use the transportation system efficiently. Agents add plans (their strategies) to their repertoire by making use of the behavioral modules. A new plan will be used immediately in order to assign a score to make it comparable to the plans already existing in the repertoire. It can be expected that the average plan score will increase during the simulation until reaching a level where the agents have found their individually best strategies.

The agents have three different possibilities to replan: route replanning, time replanning, choosing an already existing plan. As already mentioned, only a certain share of agents replan. The replanning probability is not fixed. The simulation starts with 30% of agents replanning; each of the replan options is adopted by 10% of the agents. This is a relatively high share of agents changing their behavior and by that changing the environment for the other agents as well. But this rather high replanning probability provides a quick learning process; the agents build a repertoire of plan alternatives. Later in simulation, the replanning probability is lowered to a value of 15% (5% for each replan option). This gives better average scores because of reduced fluctuations (see Fig. 2 and related text).

4 Preliminary Results

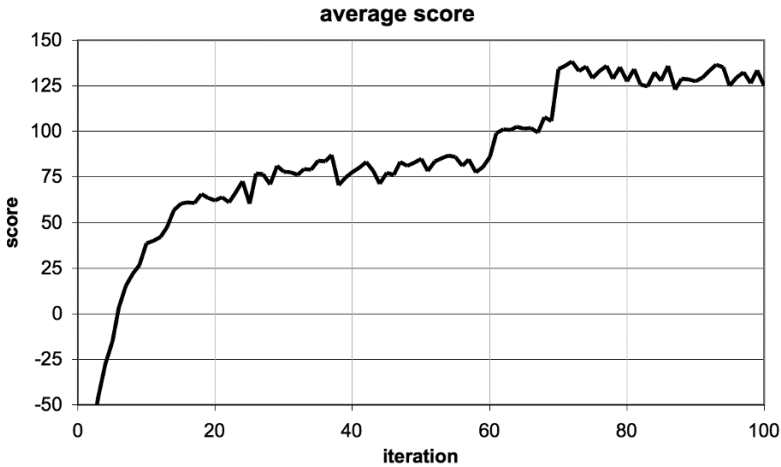


Fig. 2. Average score as a function of the iteration number

The average score gives an overview of the iterations' progress (Fig. 2). As expected, the average score is very low at the stage of initial plans, meaning the agents, in average, have not yet found good solutions for themselves. But the agents learn how to improve their situation by using different routes or changing their timing. The higher replanning probability in the beginning allows a large share of agents to learn. When the average score does not show further improvements but oscillates, the replanning probability was set to half of the original value. At iteration 60 the reconfiguration was set, which can be also seen in the figure by improving scores around iteration 60. Also with the lower replanning probability fluctuations occur, but this can be also observed in real traffic.

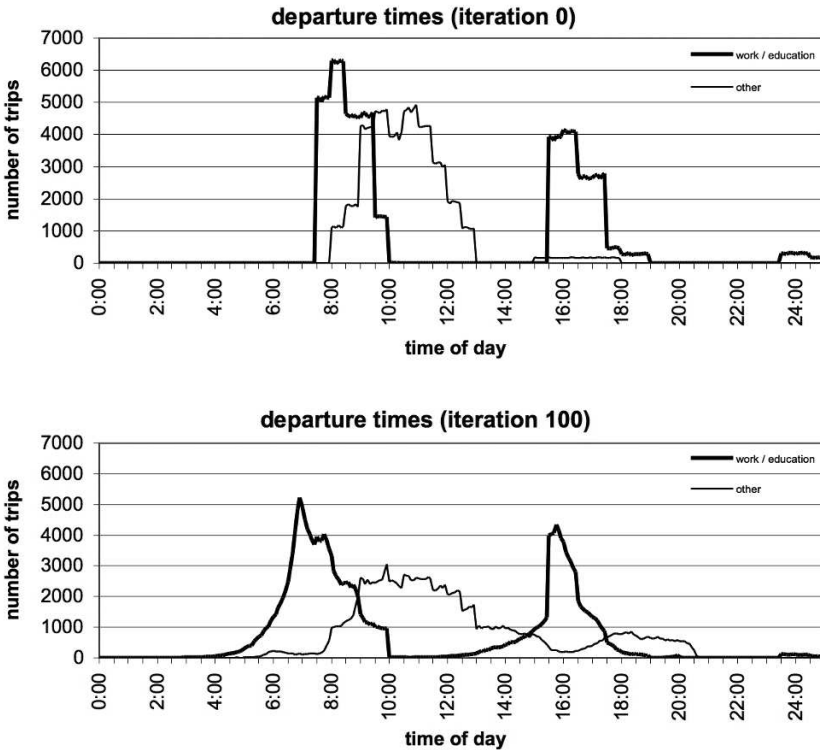


Fig. 3. Departure times distributions

Figure 3 shows the departure time distribution of two trip types: to or from work or education, and all others. The top plot shows the initial (iteration 0) departure times, based on heuristic expert knowledge, encoded in the initial conditions. The bottom plot shows the departure times after one hundred iterations. One notices the following effects:

- The initial rectangular shapes are replaced by more plausible smooth shapes.
- Travelers have, in average, moved to *earlier* departure times, with a peak before 7:00. We are, at this point, unable to judge if this is realistic.
- Those trips that are not coupled to office hours have moved to less congested time windows. Notice, in particular, the “dip” of those departures around 16:00, clearly avoiding the rush period.



Fig. 4. Daily flows of vehicles. Dark gray: less than 10,000/day; gray: around 20,000/day; black: more than 30,000/day

Figure 4 shows daily flows of vehicles. Flows below 10,000/day are displayed in dark gray, flows of 20,000/day are displayed in gray, flows above 30,000/day are displayed in black. Flows in between are displayed in interpolated colors. The figure shows the result of iteration 100. Since the simulation uses only 10% of the population, the numbers from the simulation were multiplied by 10 in order to have the same scale as real world numbers. – One observes that the pattern in the south-western sector is significantly different from the pattern in the north-eastern sector: While in the south-western half there is considerable traffic on the peripheral freeway, the patterns in the north-eastern sector are considerably more radial. This is due to extended freeway construction in the western sector during the division of the city, and the lack of such construction in the eastern sector.

It is, unfortunately, at this point difficult to say anything beyond the above. After spending considerable effort cleaning up other issues, such as related to network data or initial demand, our current issue are gridlocks as shown in Fig.5. The problem here is that all links along the loop are full, and all vehicles that are at the respective downstream ends of links want to enter the next link of the loop. In this situation, no vehicle along the loop can move, which is why it is called gridlock. Such situations can in principle occur along any closed loop of the network graph, but have a much higher probability along short loops.

We have been aware of the gridlock issue for many years, and have conventionally resolved it by the introduction of “lost vehicles”: Vehicles that could not move for a certain amount of time were taken out of the simulation. This approach, however, does not seem to be sufficient for the Berlin simulations, and we are currently investigating other solutions.

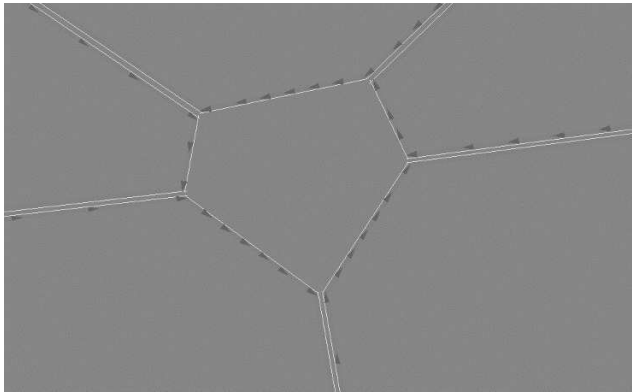


Fig. 5. Gridlock at Grosser Stern in Berlin

5 Discussion

Considerable work was necessary to adapt the network data to our purposes. Although the data requirements of the queue model are not particularly difficult (apart from the number of lanes, all information is the same as for traditional planning models), it turns out that the queue model is more sensitive to data errors than the conventional models. This is due to the hard limit on the capacity: In a conventional static assignment model, short links with reduced capacities have very little effect, whereas in the queue model, they cause large spillbacks. This effect occurs in all dynamic models with hard capacity limits. Our hope is that some of these issues will improve with the increasing availability of standardized commercially maintained network data. For the time being, however, such data are useful for routing and guidance, but do not possess reliable attributes (such as capacity) for traffic flow simulations.

Our demand generation suffers from the fact that the base model generates round trips, not daily plans. In consequence, a person who has, say, the activity chain home-work-home-leisure-home will be divided into two “virtual” agents, one with activity chain home-work-home and the other with activity chain home-leisure-home. There is no reason why those two virtual agents should perform their trips in a sequential order, so in general they will, wrongly, not do so. This issue is due to the orientation of the demand generation towards daily travel, without consideration of the time-of-day. It will probably be necessary to devise a completely new method of demand generation.

An additional problem is that our simulations currently lack commercial travel and long-distance travel. Commercial travel, in particular, increases the overall demand. Quite in general, it is difficult to get temporally consistent data – the data

that we are currently using as input comes from many different years. In most places, things do not change that quickly, and it is sufficient to have the road network data and the traffic counts from the same year. Berlin, however, is a quickly-changing city due to the re-unification, and in consequence, such differences matter considerably.

6 Summary and Conclusion

This paper provided an introduction to multi-agent traffic simulation. It included some description of where we are with respect to the implementation of a real-world Berlin scenario. It turns out that the difficulties there, because of data availability and because of the special situation of Berlin after the re-unification, are considerably larger than in previous scenarios that we have treated.

References

- Arentze T, Timmermans H (2000) Albatross: A Learning-Based Transportation Oriented Simulation System. EIRASS (European Institute of Retailing and Services Studies), TU Eindhoven, NL
- Arentze T, Timmermans H (2005) ALBATROSS – Version 2.0 – A learning based transportation oriented simulation system. EIRASS (European Institute of Retailing and Services Studies), TU Eindhoven, NL
- Arnott R, de Palma AD, Lindsey R (1993) A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review* **83**: 1:161-179
- Bando M, Hasebe K, Nakayama A, Shibata A, Sugiyama Y (1995) Dynamical model of traffic congestion and numerical simulation. *Phys Rev E* **51**: 2:1035–1042
- Beazley D, Lomdahl P, Gronbech-Jensen N, Giles R, Tamayo P (1995) Parallel algorithms for short-range molecular dynamics. In: Stauffer D (ed) *Annual reviews of computational physics III*. World Scientific, Singapore, pp 119–176.
- Beckman RJ, Baggerly KA, McKay MD (1996) Creating synthetic base-line populations. *Transportation Research Part A – Policy and Practice* **30**: 6: 415–429
- Ben-Akiva M, Lerman SR (1985) *Discrete choice analysis*. The MIT Press, Cambridge, MA
- Ben-Akiva M, Bierlaire M, Koutsopoulos H, Mishalani R (1998) Dynamit: A simulation-based system for traffic prediction. DACCORS Short Term Forecasting Workshop, The Netherlands. See also its.mit.edu
- Bhat C, Guo J, Srinivasan S, Sivakumar A (2004) A comprehensive econometric microsimulator for daily activity-travel patterns (cemdap). *Transportation Research Record* **1894**: 57–66
- Bowman J, Bradley M, Shifan Y, Lawton T, Ben-Akiva M (1999) Demonstration of an activity-based model for Portland. In *World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research 1998*, vol. 3. Elsevier, Oxford, pp 171–184

- Brockfeld E, Kühne R, Skabardonis A, Wagner P (2003) Towards a benchmarking of microscopic traffic flow models. *Transportation Research Record* **1852**: 124–129
- Cascetta E, Papola A (1998) An implicit availability/perception random utility model for path choice. In *Proceedings of TRISTAN III*, vol. 2. San Juan, Puerto Rico
- Cetin N, Burri A, Nagel K (2003) A large-scale agent-based traffic microsimulation based on queue model. In *Proceedings of Swiss Transport Research Conference (STRC)*. Monte Verita, CH. URL www.strc.ch
- Chang G, Mahmassani H, Herman R (1985) A macroparticle traffic simulation model to investigate peak-period commuter decision dynamics. *Transportation Research Record* **1005**:107–120
- Charypar D, Nagel K (2005) Q-learning for flexible learning of daily activity plans. *Transportation Research Record* **1935**: 163–169
- Chowdhury D, Santen L, Schadschneider A (2000) Statistical physics of vehicular traffic and some related systems. *Physics Reports* **329**: 4–6: 199–329
- de Palma A, Marchal F (2002) Real case applications of the fully dynamic METROPOLIS tool-box: An advocacy for large-scale mesoscopic transportation systems. *Networks and Spatial Economics* **2**: 4: 347–369
- Dijkstra E (1959) A note on two problems in connexion with graphs. *Numerische Mathematik* **1**: 269 – 271
- DYNAMIT [www page](http://www.mit.edu/its) (accessed 2005) URL mit.edu/its
- DYNASMART [www page](http://www.dynasmart.com) (accessed 2005) URL www.dynasmart.com
- Emmerink R (1996) Information and pricing in road transport. Ph.D. thesis, Tinbergen Institute, Amsterdam, NL
- Esser J (1998) Simulation von Stadtverkehr auf der Basis zellularer Automaten. Ph.D. thesis, University of Duisburg, Germany. See also www.traffic.uni-duisburg.de
- Flötteröd G, Nagel K (2005) Some practical extensions to the cell transmission model. Paper 128, IEEE Intelligent Transportation Systems Conference, Vienna
- Flötteröd G, Nagel K (2006) Modeling and estimation of combined route and activity location choice. In: IEEE Intelligent Transportation Systems Conference. Toronto, Canada, pp 394–399. URL www.itsc2006.org
- Gawron C (1998a) An iterative algorithm to determine the dynamic user equilibrium in a traffic simulation model. *International Journal of Modern Physics C* **9**: 3: 393–407
- Gawron C (1998b) Simulation-based traffic assignment. Ph.D. thesis, University of Cologne, Cologne, Germany. URL www.zaik.uni-koeln.de
- Gipps PG (1981) A behavioural car-following model for computer simulation. *Transportation Research B* **15**: 105–111
- Gloor C, Mauron L, Nagel K (2003) A pedestrian simulation for hiking in the Alps. In *Proceedings of Swiss Transport Research Conference (STRC)*. Monte Verita, CH. URL www.strc.ch
- Hall F, Agyemang-Duah K (1991) Freeway capacity drop and the definition of capacity. *Transportation Research Record* **1320**: 91–98
- Helbing D (2001) Traffic and related self-driven many-particle systems. *Reviews of Modern Physics* **73**: 1067–1141
- Helbing D, Hennecke A, Treiber M (1999) Phase diagram of traffic states in the presence of inhomogeneities. *PRL* **82**: 21: 4360–4363
- Hofbauer J, Sigmund K (1998) *Evolutionary games and replicator dynamics*. Cambridge University Press, Cambridge

- Jara-Diaz SR, Guerra R (2003) Modeling activity duration and travel choice from a common microeconomic framework. In Proceedings of the meeting of the International Association for Travel Behavior Research (IATBR). Lucerne, Switzerland. See www.ivt.baug.ethz.ch
- Jonnalagadda J N, Freedman, Davidson W, Hunt J (2001) Development of microsimulation activity-based model for San Francisco: destination and mode choice models. *Transportation Research Record* **1777**: 25–35
- Jost D, Nagel K (2003) Probabilistic traffic flow breakdown in stochastic car following models. *Transportation Research Record* **1852**: 152–158
- Kerner B, Klenov S, Wolf D (2002) Cellular automata approach to three-phase traffic theory. *J Physics A: Math General* **35**: 47: 9971–10013
- Kerner BS (1998) Experimental features of self-organization in traffic flow. *Physical Review Letters* **81**: 3797–3800
- Krauß S, Wagner P, Gawron C (1996) Continuous limit of the Nagel-Schreckenberg model. *Phys Rev E* **54**: 4: 3707–3712
- Kutter E (1984) Integrierte Berechnung städtischen Personenverkehrs – Dokumentation der Entwicklung eines Verkehrsberechnungsmodells für die Verkehrsentwicklungsplanung Berlin (West). Arbeitsberichte zur Integrierten Verkehrsplanung, TU Berlin, Berlin
- Kutter E, Mikota HJ (1990) Weiterentwicklung des Personenverkehrsmodells Berlin auf der Basis der Verkehrsentstehungsmatrix 1986 (BVG). Tech. rep., Untersuchung im Auftrag des Senators für Arbeit, Verkehr und Betriebe, Berlin
- Kutter E, Mikota HJ, Rügenapp J, Steinmeyer I (2002) Untersuchung auf der Basis der Haushaltsbefragung 1998 (Berlin und Umland) zur Aktualisierung des Modells “Pers Verk Berlin / RPlan”, sowie speziell der Entwicklung der Verhaltensparameter ’86–’98 im Westteil Berlins, der Validierung bisheriger Hypothesen zum Verhalten im Ostteil, der Bestimmung von Verhaltensparametern für das Umland. Draft of the final report, Sponsored by the “Senatsverwaltung für Stadtentwicklung Berlin”, Berlin/Hamburg
- Lighthill MJ, Whitham JB (1955) On kinematic waves. I: Flow movement in long rivers. II: A Theory of traffic flow on long crowded roads. *Proceedings of the Royal Society A* **229**: 281–345
- MATSIM [www page](http://www.matsim.org) (accessed 2005) MultiAgent Transportation SIMulation. URL www.matsim.org
- Meister K, Balmer M, Axhausen K, Nagel K (2006) planomat: A comprehensive scheduler for a large-scale multi-agent transportation simulation. In Proceedings of the meeting of the International Association for Travel Behavior Research (IATBR). Kyoto, Japan. See www.iatbr.org
- Miller E, Roorda M (2003) A prototype model of household activity/travel scheduling. *Transportation Research Record* **1831**: 114–121
- Nagel K, Nelson P (2005) A critical comparison of the kinematic-wave model with observational data. In: Mahmassani H (ed) *Transportation and Traffic Theory – Flow, Dynamics and Human Interaction*. Elsevier, Amsterdam, pp 145–164
- Nagel K, Rickert M (2001) Parallel implementation of the TRANSIMS micro-simulation. *Parallel Computing* **27**: 12: 1611–1639
- Pendyala R (accessed 2005) Phased implementation of a multimodal activity-based travel demand modeling system in florida. URL www.dot.state.fl.us/research-center/Completed_PTO.htm. Project number 0510812 (BA496)
- Rügenapp J, Steinmeyer I (2006) Activity-based demand generation: Anwendung des Berliner Personenverkehrsmodells zur Erzeugung von Aktivitätsketten als Input für

- Multi-Agenten-Simulationen. Working Paper 06-09, Fachgebiet Verkehrssystemplanung und Verkehrstelematik, TU Berlin, Berlin, Germany. URL www.vsp.tu-berlin.de
- Salvini P, Miller E (2005) ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics* **5**: 2: 217–234
- Schuster HG (1995) *Deterministic Chaos: An Introduction*. Wiley-VCH Verlag GmbH
- Stauffer D (1991) Computer simulations of cellular automata. *J Phys A* **24**: 909–927
- Strauch D, Hertkorn G, Wagner P, Kühne R (2002) Neue Ansätze zu einer mikroskopisch-dynamischen Verkehrs- und Flächennutzungsplanung im verbundprojekt ILUMASS. In: Tavangarian D, Grützner R (eds) *Simulationstechnik, Frontiers in Simulation*, vol 12, pp 511–516. SCS Publishing House
- SUMO [www page](http://www.sumo.sourceforge.org) (accessed 2005) URL sumo.sourceforge.org
- TRANSIMS [www page](http://www.transims.tsasa.lanl.gov) (accessed 2005) TRansportation ANalysis and SIMulation System. URL transims.tsasa.lanl.gov. Los Alamos National Laboratory, Los Alamos, NM
- Waddell P, Borning A, Noth M, Freier N, Becke M, Ulfarsson G (2003) Microsimulation of urban development and location choices: Design and implementation of UrbanSim. *Networks and Spatial Economics* **3**: 1: 43–67

Hybrid Geographical Models of Urban Spatial Structure and Behaviour

Mark Birkin

School of Geography, University of Leeds, UK

Abstract

The chapter explores the introduction of simple behavioural mechanisms into a spatial microsimulation model. It is hypothesised that the representation of individual behaviour using appropriate rules can allow the outcome from meso-level models, such as spatial interaction models, to be reproduced. The dynamic properties of these models are also explored. Therefore the research seeks to establish links between microsimulation, agent-based approaches and spatial interaction models, at a variety of scales. The simulations are operationalised in the context of a British city.

1 Introduction

From primarily economic origins, the use of microsimulation models has become popular amongst geographers and regional scientists in recent years. Microsimulation models allow a rich representation of the attributes and spatial location of a population. They are particularly useful in supporting the adoption of rule-based protocols (for example, ageing or eligibility for state benefits) and for dynamic updating.

More recently, these interests have been outstripped by a new focus on agent-based simulation. This technique has enjoyed a surge in popularity in both the social and biological sciences. The boundary between agent-based methods and microsimulation is not entirely clear. As a working assumption, this paper adopts the principle that microsimulation is a technique in which independence between individuals is retained, whereas agent-based methods are predicated on the representation of interaction and interdependence between the individual agents. It may also be argued that the representation of (autonomous) behaviour of individuals is a central feature of agent-based approaches. One of the objectives of this paper will be to explore the addition of a behavioural mechanism to a spatial microsimulation model.

A common feature of spatial modelling approaches at the individual level, whether implemented through microsimulation or agents, is that they are concerned with 'fast dynamics', that is to say patterns of response to a fixed backcloth or infrastructure. A typical example of this might be a traffic simulation, which represents congestion on a road network, and possible adjustments to this in terms

of changed signalling or the reconfiguration of junctions. In this example, the most important ‘slow dynamic’ would be the development of the road network itself. In this paper, we will use examples from the housing and employment system. The fast dynamic assumes that house prices are fixed; the slow dynamic includes prices as a market-balancing mechanism. In other words, the paper is concerned to develop an approach that integrates individual and market level behaviours (see Fig. 1).



Fig. 1. Left: Traditional approach: individuals respond to an infrastructure which is fixed. Right: approach proposed in this paper: mutual dependance between behaviour and infrastructure

The underlying rationale for this paper is a belief in the applicability of simulation as a planning tool. The analogy with the ‘SimCity’TM computer game is a powerful one. The simulation environment provides a tremendously rich planning framework in which gamers can refine their skills by designing ever more complex and harmonious cityscapes. Given the tremendously rich databases and powerful computational environments which are now available to city planners, why are the equivalent technologies not able to offer a similar capability in the ‘real world’? One of the key ingredients of ‘SimCity for Real’ would be an ability to relate individual behaviour to structural change over time.

The paper will be substantiated through some simple simulations, which are based on the city of Leeds in northern England, using data from the 2001 UK Census of Population and Households. First, however, we discuss some of the most important literature which is relevant to this research.

2 Microsimulation and Spatial Evolution

Microsimulation approaches originated as a method for analysing the impact of financial policy across the population. From the 1970s onwards, the technique has been embraced with enthusiasm by geographers – its advantages include the ability to link together databases, and to represent a rich array of demographic, economic and locational characteristics of the population (e.g. Birkin and Clarke

1988) and to simulate dynamic changes for the purpose of updating or forecasting small area populations (Orcutt et al. 1986). Good recent examples of applied microsimulation models are the SimBritain and WaND projects. SimBritain aims to envision the future social geography of the country by combining detailed demographics from the census with rich but spatially aggregate data from the British Household Panel Survey (Ballas et al. 2005). WaND combines detailed household data and projections with patterns of domestic water usage in order to forecast national and local water resource requirements deep into the twenty-first century (Rees 2005).

An early attempt to enhance microsimulation models by the addition of behaviour relating to activities such as retailing and the search for jobs can be found in the work of Birkin and Clarke (1987). The authors coupled a microsimulation model of individual and household demographics together with a spatial interaction model of retail flows. The spatial interaction model was used to generate interaction probabilities for individual customer groups, and on this basis individual retail destinations were then added to the micro-data set as a new characteristic. Although elements of this approach have been revisited by Nakaya et al (2003), there has been no refinement or redevelopment of this work in the intervening period.

The link between spatial interaction models and spatial structure was explored in a classic paper by Harris and Wilson (1978), and further elaborated by Clarke and Wilson (1983). A means for the introduction of new variables such as urban land prices into such models was described by Birkin and Wilson (1989). In summary, therefore, there exists a group of methods in this work which seems capable of connecting spatial micro-data to meso-level activities and interactions; and for inferring macro-level structures and patterns from these meso-level interactions. Thus it is our contention that the connection between microsimulation and spatial interaction models remains important as a mean to unlock the complex interdependencies between individual behaviour and the evolution of urban spatial structure.

A potential objection to this line of research is that other methods have now been developed which offer more promise. The most obvious contenders are agent-based models, which offer an alternative and behaviourally explicit approach to modelling activities at the individual level. Such methods have been applied with interesting results in areas such as transportation planning (Nagel 2003), anthropology (Murphy 2003), and economics (Gjerstadt and Dickhaut 1998) as well as geography. One particularly interesting class of applications is to problems of market trading (Cliff 2003), since this looks like a classic micro-macro problem, in which prices in a market adjust and balance to individual demand schedules. However in an applied urban modelling context focussing on the market for retail petroleum, Heppenstall et al. (2005) reported relatively poor performance from agent-based methods and found vastly superior results from coupling an agent model of petrol stations with a spatial interaction model of consumer behaviour. We assert that combining an individual level model with a spatial interaction model makes sense simply because spatial modelling environments are highly complex. For example, the trading simulations discussed above

typically assume a small number of buyers and sellers ($n < 100$) and these will just not scale up to real geographical problems. This is not a problem which is uniquely geographical. It seems equally unlikely that the Treasury's national economic model will be substituted for an agent-based alternative in the near future for similar reasons.

As a starting point for the research to be reported in this paper, we begin from a different perspective on the problem of linkage between models at the micro- and meso-scale. On the basis of work which again originates in the 1970s and 1980s, we specify a model at the micro-scale which is known to exhibit properties at the meso-scale which are similar to a class of spatial interaction models. We use an analysis in the style of Harris-Wilson to pursue this equivalence further, and from there we move on to propose a more complex model with appealing and useful properties.

3 Simple Models of the Urban Labour Market

3.1 Doubly-Constrained Spatial Interaction Model

We begin with a straightforward model which is capable of allocating journey-to-work flows between zones of known employment and zones of known residence. For the purposes of the numerical work to be reported in this paper, we will make the simplifying assumption throughout that each household contains one and only one economically active resident, and that full employment is maintained. No flows are allowed across the boundary into or out of the city region.

In order to generate the data for this section, journey-to-work data for the city of Leeds was extracted from the 2001 census and aggregated by census ward. There are 33 such wards in the city. The row and column totals are generated in the obvious way:

$$H_i(obs) = \sum_j T_{ij}(obs) \quad (1)$$

$$E_j(obs) = \sum_i T_{ij}(obs) \quad (2)$$

Estimated flows can be generated from a doubly-constrained spatial interaction model as follows:

$$T_{ij} = A_i B_j H_i(obs) E_j(obs) \exp(-\beta_{C_{ij}}) \quad (3)$$

$$A_i = 1 / \sum_j B_j E_j(obs) \exp(-\beta_{C_{ij}}) \quad (4)$$

$$B_j = 1 / \sum_i A_i H_i(obs) \exp(-\beta_{C_{ij}}) \quad (5)$$

By convention in this paper, we will show observed values through the parenthesis (obs). Values not marked in this way are modelled.

Since the balancing factors A_i and B_j are interdependent, the model equations need to be solved iteratively. This procedure is typically efficient and well-behaved.

The level of dispersion in the model is controlled through the distance deterrence parameter β - low values of β generate long trips and high dispersion; high values of β generate shorter trips and in the limit the model reproduces a linear programming solution with minimum trip costs (Evans 1973). Some examples are shown for this application in Fig. 2. The deterrence parameter can be calibrated to match the trip length in the observed data set. In this example, the mean trip length from the census data is 5.14 km, and this is reflected in a value of $\beta = 0.2985$. The goodness of fit between the modelled and observed flows is illustrated in Fig. 3. This model fits the data with an r-squared value of 0.85.

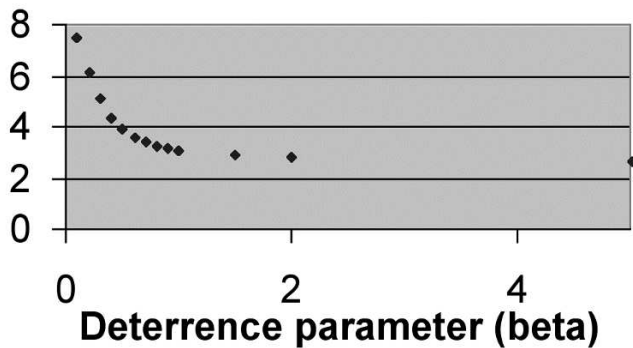


Fig. 2. Sample calibration statistics for DCSIM

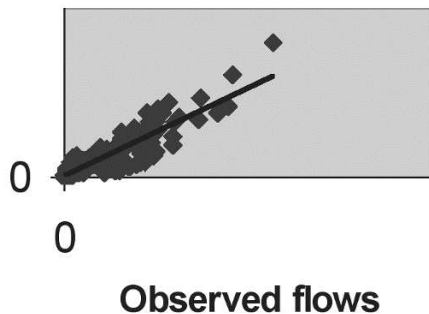


Fig. 3. Performance of the DCSIM

3.2 Micro-Simulation Model of Journey-to-Work Flows

A different approach to the problem of trip choice is offered by random utility theory. At its simplest, we can conceive of workers at employment destinations (j) searching for housing between locations (i) in order to maximise the 'utility' associated with each destination. The obvious component of this utility is the trip cost between i and j ; the less obvious element is a random component which reflects some combination of imperfect information, sub-optimal decision-making and utilities which are not observed or allowed for in the model. Thus for each employee, we can establish a utility associated with residential location i :

$$u_{ji} = \varepsilon_{ij} - c_{ij} \quad (6)$$

One of the most interesting features of this model is that under certain assumptions about the distribution of the random utility terms ε , the model is equivalent to the multinomial logit model:

$$p_{ij} = \exp(-\lambda_{C_{ij}}) / \sum_i \exp(-\lambda_{C_{ij}}) \quad (7)$$

The precise conditions for this equivalence are that the error terms should take the form of a Weibull distribution with a standard deviation of $\Pi / (\lambda\sqrt{6})$ and a mean of c' (Williams 1977). Similarities to the spatial interaction models introduced above are obvious and well-known.

In order to test the implications further, we developed a simulation as follows. We assumed a workforce of 100,000 employees, distributed across the city in accordance with the real population. Employees are selected at random, and on the basis of the workplace, each residence is assigned a utility with two components described in Eq. 6 above. The random utility terms are assigned using a Monte Carlo process from a normal distribution with mean x and standard deviation u . x and u are parameters of the simulation. For each residence, we test that housing is available at that location. Then a location is selected which yields maximum utility. The process is repeated for all employees, as illustrated in Fig. 4.

The next step is to calibrate the simulation model against the random utility parameters x and u . It turns out that variations in the mean have no impact on trip distances in the model: this appears to bear out the mathematical expectation that "the resultant demand model(s) are ... invariant with respect to the addition or subtraction of constants ... and the mean value may effectively be ignored" (Williams 1977). The standard deviation of the utility component is, however, monotonically related to the model trip length as shown in Table 1. When the standard deviation is set to 5.65, the observed mean trip length of 5.14 km is reproduced. The scatter of observed versus predicted interactions yields a correlation r -squared = 0.83. The fit of the random utility model versus the DCSIM has r -squared = 0.94.

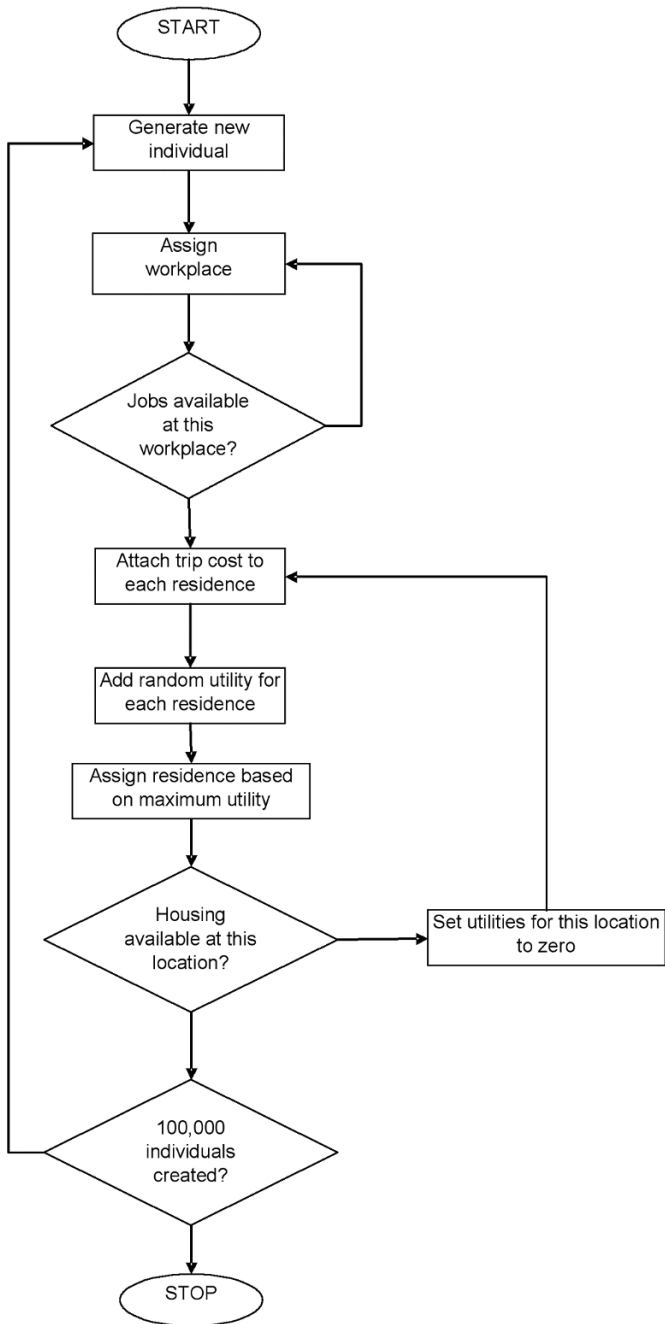


Fig. 4. Algorithm to generate individuals with best workplace-residence location

The interpretation and hypothesis associated with these results is that the microsimulation model is slightly less effective than the DCSIM and that this could be due to the operation of the origin (residence) constraints within the two models. In the DCSIM, the availability of housing operates as a ‘balancing factor’ which will tend to dampen down flows to the most accessible destinations; in the simulation model, accessible destinations will simply fill up first, and then overspill into neighbouring zones. If this interpretation is correct then one would expect the fit between the DCSIM and the simulation model to deteriorate as the linear programming solution is approached for high values of beta (and low values of u in the microsimulation model). However initial experiments are inconclusive: for $u = 1$, $\beta = 0.9$ both cases yield an average distance of 3.22 km, but the cross-model goodness-of-fit again registers $r\text{-squared} = 0.94$.

We now proceed as follows. Given the perceived asymmetry between the DCSIM and the intrinsically singly-constrained nature of the microsimulation (logit-analogue) model, we seek to respecify the DCSIM as an equilibrium solution to a singly-constrained model, and use this as a basis for revising the microsimulation model in an equivalent form.

4 Semi-Aggregate Models of the Urban Labour Market

4.1 Quasi-Doubly Constrained Model

Consider the following singly-constrained spatial interaction model:

$$T_{ij} = B_j W_i E_j(\text{obs}) \exp(-\beta_{C_{ij}}) \quad (8)$$

$$B_j = 1 / \sum_i W_i \exp(-\beta_{C_{ij}}) \quad (9)$$

W_i can be considered as a representation of the ‘attractiveness’ of each residential zone i . By analogy with the argument of Harris and Wilson (1978) we can argue that the attractiveness of residential locations can be ‘evolved’ over ‘time’ in accordance with the known distribution of housing, H_i . Thus:

$$\Delta W_i = \varepsilon \left(\sum_j T_{ij} - H_i \right) \quad (10)$$

The equilibrium conditions to this adjustment process are clearly equivalent to the origin constraints of the DCSIM introduced in Section 3 above. However the equivalence between the two models is hardly trivial. In computational terms, an alternate adjustment mechanism in the balancing factors A_i and B_j is replaced by an embedded iterative scheme in which the destination balancing factors are solved at each increment within the origin attractiveness terms. Nevertheless the two models can be shown to be equivalent computationally for a range of deterrence factors (see Table 1). Two pertinent questions relating to the attractiveness terms are now as follows:

1. Can we use the attractiveness terms to respecify the utility component of the microsimulation model in such a way as to demonstrate equivalence to the quasi-doubly constrained spatial interaction model?
2. Do the attractiveness terms bear interpretation in terms of the characteristics of the associated residential locations?

4.2 Micro-Simulation Model

Theory suggests that a suitable strategy would be the replacement of the cost component to utility with a ‘consumers’ surplus’ for each individual in the form:

$$CS = \log (W_i) - \beta_{C_y} + \lambda \tag{11}$$

In our next experiment, we assume that the attractiveness components of consumers’ surplus, and the distance deterrence parameter beta, can both be translated directly from the spatial interaction model from Section 4.1 above. These results are summarised in Table 1 and show that the situation is actually quite complex, since now the parametrisation of beta against average distance travelled does not yield the best fit to the interaction pattern between zones. This suggests either:

- a) that the subtleties regarding the distribution of utilities (see discussion of Section 3.2 above) requires further investigation; or
- b) that the ‘maximum likelihood’ (average distance) solution is not necessarily a statistical optimum.

Table 1. Outputs statistics for the agent-based model

Av dist	0.1	0.2	0.3	0.4	0.5
0	5.55	3.97	3.07	3.07	
1	6.79	5	3.8	3.53	3.41
2	7.56	5.98	4.95	4.44	4.16
3	7.83	6.68	5.78	5.29	4.95
GOF	0.1	0.2	0.3	0.4	0.5
0	0.3927	0.4044	0.541	0.6604	
1	0.6356	0.7878	0.7967	0.8117	0.8229
2	0.5267	0.6835	0.7769	0.8256	0.8523
3	0.469	0.5835	0.7208	0.7921	0.8257
	Beta	StDev	AvDist	Entropy	R2(obs)
	0.2985	2.2	5.15	89.93	0.7636

Av dist – Average distance

GOF – Goodness of fit (r-squared)

0.1– 0.5 : values of beta

0,1,2,3 : values of lambda

These are shown in Fig. 5, which makes a lot of sense in terms of the social geography of Leeds. It is also possible to try and model the attractiveness terms in

terms of area characteristics like house prices, crime and quality of local services, like schools.



Fig. 5. Attractiveness by ward QDCSIM

5 An Interacting Fields Model

Ira Lowry's Model of Metropolis (1964) is a classic of the modelling genre of the 1950s and 1960s, in which the author seeks to build an 'instant city' using only a distribution of primary (or 'basic') employment as a building block. On the basis of these initial jobs, workers are assigned to residences, and on the basis of residential demand for services, retail centres spring up in the city, generating further employment within the service sector. Once these interactions between basic employment, residential location and service sector employment are stabilised then the resulting city structure can be analysed. The elegance of Lowry's model has been widely admired and influenced a generation of land use and transportation modelling approaches in the US and elsewhere (e.g. Putnam 1974; Webber 1980; Foot 1981).

Lowry's original model was calibrated to the city of Pittsburgh. In this paper we redevelop the model for the city of Leeds in the spirit of the original implementation. Thus Lowry divides the city of Pittsburgh into uniform tracts of one square mile. Here we adopt a similar grid of kilometre squares. As Lowry identifies three categories of retail and service employment (labelled as 'neighbourhood', 'local' and 'metropolitan'), so three similar categories are identified here.

Lowry's workforce is segmented into four occupation groups, and again a similar typology is followed.

However the implementation of the model is distinctive in two important ways. Firstly, the apparatus of sections 3 and 4 above provides the means to implement the interacting fields concept using a population of individual agents in preference to the usual array-based representation of population cohorts. Secondly, the Harris-Wilson mechanism for spatial dynamics provides a more flexible means to accommodate the development of residential and retail structure (it is hard to escape the feeling on re-reading Lowry's original treatise that the retail distributions in particular are strongly constrained by artificial minimum and maximum size thresholds rather than being distilled from some more fundamental centralising process).

Our modelling strategy is therefore as follows. To begin with, we establish a lattice of grid-squares to represent the city of Leeds. The city is bounded by a rectangle 33 kilometres wide by 25 km high. Of the 825 cells, 424 are outside the city boundary and the remainder within. For each of the 401 internal grid cells, we establish a level of basic employment from the 2001 Annual Census of Employment. This gives 229,850 jobs, with another 170,600 retail and service jobs to be allocated within the model.

Each of the 229,850 employees is created as an agent with one of the four occupation types. These agents must now search for somewhere to live. The residential choice process is governed by a utility function which trades off accessibility against density of land use. The blue collar workers tend to place a higher value on accessibility; the white collar workers tend to favour low density housing. This gives rise to a broadly concentric pattern of residential land use around the major employment centres.

Next, retail demand is generated by residents. Expenditure is distributed between retail centres which are initially assumed to be uniformly distributed across the city. The three different activity types (neighbourhood, local and metropolitan) are governed by varying levels of attractiveness and distance deterrence. Each centre is governed by an evolutionary dynamic in relation to its ability to attract customers and the likely costs of floorspace provision. In this way, a smaller number of significant centres is allowed to grow and develop. The process of retail centre evolution is embedded within the residential location algorithm, so that a stable set of centres is allowed to emerge before the next phase, in which further agents are created in order to meet the employment needs of retail and service businesses. The new agents then need to enter the 'housing market' and the cycle begins again.

Some results from the simulations are shown in Figs. 6 and 7. Figure 6 shows the distribution of housing for employees in four categories – respectively Social Class AB, C1, C2 and DE. The overall texture of the maps indicates that the lower social classes have higher densities of occupation, as one would expect. The main high social grade zone is to the north-west of the city. Other suburbs and surrounding locations, such as Wetherby to the north east, are mainly colonised by social grades C, D and E. It is tempting to suggest that this pattern is more representative of Victorian Leeds than the present day city – further exploration of

journey-to-work data and model calibrations are required to investigate this hypothesis. Figure 7 shows the distribution of the three retail types of neighbourhood, local and metropolitan. Here we see the development of a single metropolitan, albeit it dispersed across a number of central city zones. Local retail centres are spread against five substantial locations, while neighbourhood retail activity is the most evenly dispersed: all in accordance with our expectations.

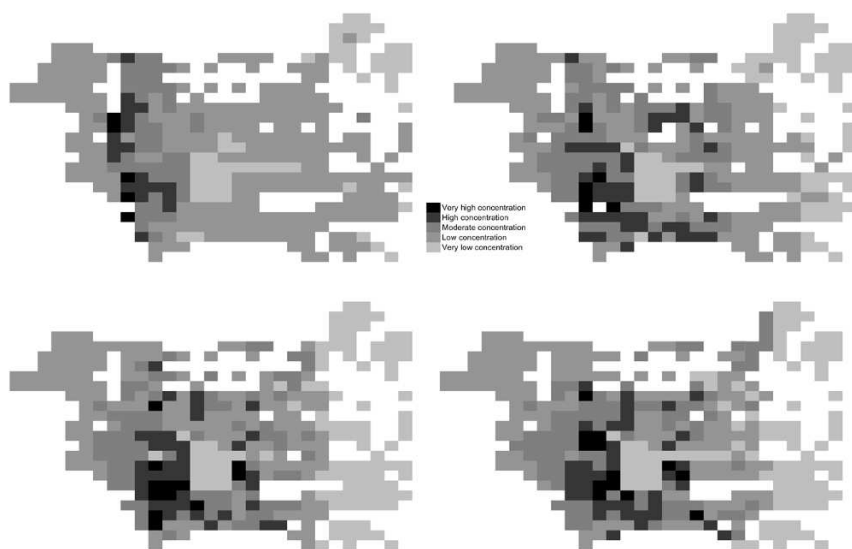


Fig. 6. Distribution of the simulated housing for employees in four categories: unskilled workers (above left); semi-skilled workers (above right); skilled manual workers (below left); professional workers and managers (below right).



Fig. 7. Distribution of three retail floorspace types: neighbourhood (top), local (centre) and metropolitan (bottom)

6 Conclusions

One of the main conclusions is that the equivalence between spatial interaction models and microsimulation is not straightforward, even for relatively simple

specifications. Practical applications demand highly disaggregate models, which will be even less tractable.

The suggestion of the paper is that parameters are not translatable between scales, and this is not entirely surprising. This could also imply that there are dangers in translating interaction probabilities directly from the meso- to the micro-scale and back again.

One other possibility is perhaps to estimate utilities directly at the micro-scale, and then where meso-scale spatial models are needed to fit them to the micro-simulation data. This approach is perhaps worthy of further investigation.

References

- Ballas D, Rossiter D, Thomas B, Clarke GP, Dorling D (2005) *Geography matters: simulating the local impacts of national social policies*. Joseph Rowntree Foundation, York (forthcoming)
- Birkin M, Clarke M (1987) Comprehensive models and efficient accounting frameworks for urban and regional systems. In: Griffith D, Haining R (eds) *Transformations through space and time*. Martinus Nijhoff, The Hague, pp 169-195
- Birkin M, Clarke M (1988) SYNTHESIS: A SYNTHetic Spatial Information System for urban modelling and spatial planning. *Environment and Planning A* **20**: 1645-1671
- Birkin M, Clarke M (1989) The generation of individual and household incomes at the small area level using Synthesis. *Regional Studies* **23**: 535-548
- Birkin M, Wilson AG (1989) Some properties of spatial- structural- economic dynamic urban models. In: Hauer J, Timmermans H, Wrigley N (eds) *Progress in the dynamic analysis of spatial systems*. Reidel, pp 184-200
- Clarke M, Wilson AG (1983) The dynamics of urban spatial structure: Progress and problems. *Journal of Regional Science* **23**: 1-18
- Cliff D (2003) Explorations in evolutionary design of online auction market mechanisms. *Journal of Electronic Commerce Research and Applications* **2**: 2: 162-175
- Evans S (1973) A relationship between the gravity model for trip distribution and the transportation problem in linear programming. *Transportation Research* **7**: 39-61
- Foot D (1981) *Urban models*. Methuen, London
- Gilbert N (2003) *Multi-agent based simulation*. ESRC Research Methods Workshop, Leeds
- Gjerstad S, Dickhaut J (1998) Price formation in double auctions. *Games and Economic Behaviour* **22**: 1: 1-29
- Harris B, Wilson AG (1978) Equilibrium values and dynamics of attractiveness terms in production-constrained spatial interaction models. *Environment and Planning A* **10**: 371-388
- Heppenstall A, Evans A, Birkin M (2005) A hybrid multi-agent/spatial interaction model system for petrol price setting. *Transactions in GIS* **9**: 35-51
- Lowry I (1964) *A Model of Metropolis*. The Rand Corporation, Santa Monica, CA
- Murphy M (2003) Including feedback in demographic microsimulation models. ESRC Research Methods Workshop, Leeds
- Nagel K (2003) Traffic networks. In: Bornholdt S, Schuster H (eds) *Handbook of Graphs and Networks: From the Genome to the Internet*. Wiley, New York, pp 248-272

- Nakaya T, Yano K, Fotheringham AS, Ballas D, Clarke GP (2003) Retail interaction modelling using meso and micro approaches. Paper presented at the 33 rd Regional Science Association, RSAI – British and Irish Section conference, St Andrews, Scotland, August
- Orcutt GH, Merz J, Quinke H (eds) (1986) *Microanalytic Simulation Models to Support Social and Financial Policy*. North-Holland, Amsterdam
- Putnam S (1974) Preliminary results from an integrated transportation and land-use models package. *Transportation* **3**: 193-224
- Rees PH (2005) *Subnational Population Projections in the United Kingdom: Progress Report*. Seminar presented at the Institute of Geography and Spatial Organisation, Polish Academy of Sciences, Warsaw
- Webber M (1980) *Information theory and urban spatial structure*. Croom Helm, London
- Williams HCWL (1977) On the formation of travel demand models and economic evaluation of user benefit. *Environment and Planning A* **9**: 285-344
- Zedlewski S (1990) The development of the Dynamic Simulation of Income Model (DYNASIM). In: Lewis G, Michel S (eds) *Microsimulation Techniques for Tax and Transfer Analysis*. The Urban Institute Press, Washington DC, pp 109-136

Two Complexities and a Few Models

Ivan Blečić, Arnaldo Cecchini, Giuseppe A. Trunfio

Laboratory of Analysis and Models for Planning (LAMP), University of Sassari, Italy

Abstract

The difficulty in dealing with urban systems' complexity and the related difficulty to analyse and forecast is twofold: one kind of difficulty lies in the complexity of the system itself, and the other is due to the actions of actors, which are "acts of freedom". In our contribution we would like to present a set of techniques and models, with respective software packages (MaGIA, The Time Machine, *CAGE* and *GioCoMo*), that have proven to be of great potential for enactment and management of communication, participatory, consensus-building and simulation processes. As such, our approach tries to cope with both aspects of complexity mentioned above.

1 Premises

May we be allowed to make a healthful excursion into the profession of planner, and into that of his natural ally the analyst, which includes the subspecies of the modeller¹. Hopefully, this will help to explain better why we have founded a Laboratory of Analysis and Models for Planning.

Let us begin with a set of schemes which in some way characterise the relationship between planner and those who commission work to him or her, which in general is a public subject representing "general interests", in one way or another.

Scheme 1.

Client ←—————→ Expert

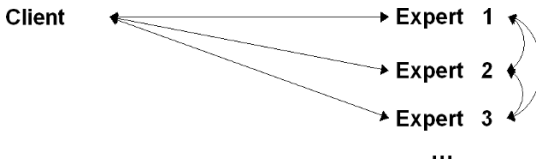
This is a classical scheme; it is the scheme Simon (1988) is referring to when he asks the question "who is the client?"; Simon holds that this question can in some cases (those he calls "microsocial") have a "simple" answer:

¹ The schemes we will be referring to in this text are appropriate for the organisational, political and cultural model in Italy and generally in the European Union, and cannot be mechanically transposed to different contexts. Notwithstanding this, probably many of the problems outlined here are independent of the socio-political context.

It may seem peculiar to ask, “Who is the client?” when speaking of the design of large social systems. The question need not be raised about smaller-scale design tasks, since the answer is built into the definition of the professional roles of designers. At microsocial levels of design it is tacitly assumed that the professional architect, attorney, civil engineer, or physician works for a specific client and that the needs and wishes of the client determine the goals of the professional’s work.” (Simon 1981, p. 173)

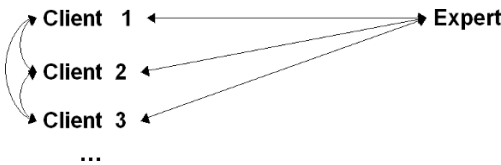
Notwithstanding its simplicity, the domain and the modality of application of this scheme are however less trivial than might be thought. The relationship between a client and a professional is frequently quite complicated and sometimes the professional does not mind being his/her own client, just as the client does not mind interfering with the professional. Suffice it to recall the relationship between the architect, the designer of a house and his/her client, the future house owner; or in a grand way, the relationship between Michelangelo and Pope Julius II. Due to increasingly pervasive and comprehensive division and specialisation of labour in the modern world, the scheme will more likely tend to resemble to something like this:

Scheme 2.

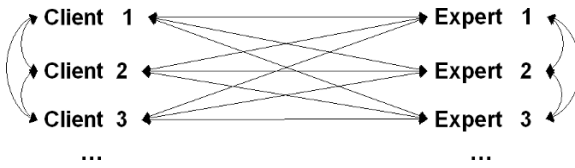


It is a scheme that brings about a few additional problems. Even allowing for a hierarchy among experts, the number of relations a client has to activate are multiple. Also, experts are generally experts in different fields and therefore have different priorities, values, languages and objectives. A variation of the scheme might be the following:

Scheme 3.



Another complex scheme, even if there were a hierarchy among clients (e.g. had the clients been the couple and the expert the architect, it should not been taken for granted that clients share the same visions and desires); in reality, much more frequently, the situation is:

Scheme 4.

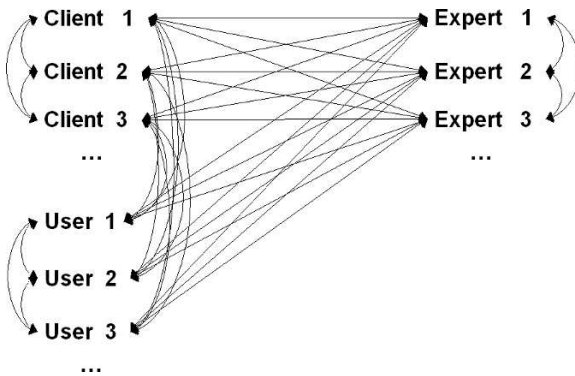
This one sums up all the previous complications and adds a few more, as may be self-evident. And on top of all that, Simon himself affirms that these schemes are actually not good enough in some cases, and are definitely not good enough in the case of the planner. In fact:

The architect need not decide if the funds his client wants to spend for a house would be better spent, from society's standpoint, on housing for low-income families. The physician need not ask whether society would be better off if his patient were dead.

[...] But as knowledge grows, the role of the professional comes under questioning. [...]

Whether through the modification of professional norms or through direct intervention of government, new obligations are placed on the professional to take account of the external effects - the consequence beyond the client's concern - that are produced by the design." (Simon 1981, pp. 173-174)

Altogether, the situations described are still quite optimistic for the case of the actual work of a planner when compared to an architect: in fact, in the planner's case, the user or – to be brief – the users appears thus:

Scheme 5.

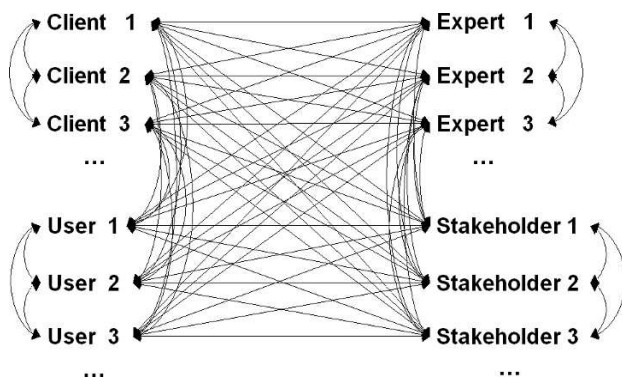
In point of fact – as we have said before – normally a planner's clients are public authorities. We can even put aside the fact that these are in general rather articulated organisational structures, with an internal diversity of objectives: as is for instance the case, not only in Italy, between "technicians" and "politicians" in a Public Administration, the former generally being a permanent presence, while the latter "lasting" only 5-10 years for the duration of their political mandate. So even putting aside such internal complexity, a public administration in principle expresses the interests of users (of citizens, and their electorate), but never actually coincides exactly with them. There are a number of reasons to explain why the

clients, those exercising a representative function, are not necessarily or completely the expression of the users: for example, some users might not be citizens and voters *tout court* (like for instance persons under the legal age), or might not be so in that country (foreigners), or in that particular administrative area (commuters and non-residential city users). In other cases, some users might not share the opinions of the public administration; and in political systems such as a representative democracy, with representatives not being their proxies, these users could even be the majority. Or users might not share their representatives' opinion for other reasons, for example maintaining that a problem is no longer important or preferring other solutions instead of those proffered at the moment they cast their vote. Relations are numerous, complicated and difficult.

“When planning was first institutionalised as part of modern western democracies, the process was entirely technical, operated by architects supported through philanthropists and reformers but implemented as top-down *fait accompli*. This model was only rarely modified for it was simply assumed that insofar as the public had any relevance to the process of preparing plans, this was purely to inform and educate, rather than involve in any direct way. Patrick Geddes' (1915, 1949) civic exhibitions were the nearest that the process ever came to participation [...]. For the next 60 years, the process of involving groups other than the experts who prepared the plan slowly widened. [...] This process had come to be known as public participation by the 1960s but it was conceived as something which was tacked onto the technical process of plan preparation rather than as an activity which suggested involvement of the wider public in the actual process of plan making.” (Batty 1997, p. 17)

Communication problems are many and various. But does it end here?

Scheme 6.



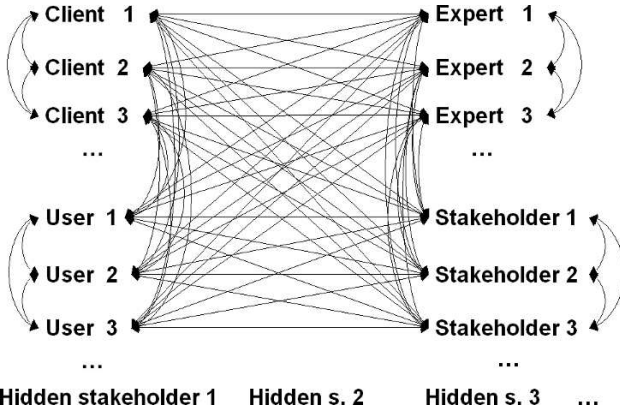
That's right! There are also the interest groups, many of which may not even be users. Suffice to think of the developer who owns the land or is financing the development, or the construction workers: As Snoopy would have it: “the plot thickens...”.

But the fog is never thick enough!

The idea that all the interest groups are explicit and declared “actors” is rather extravagant (in years of experience in gaming-simulation, we have never seen a

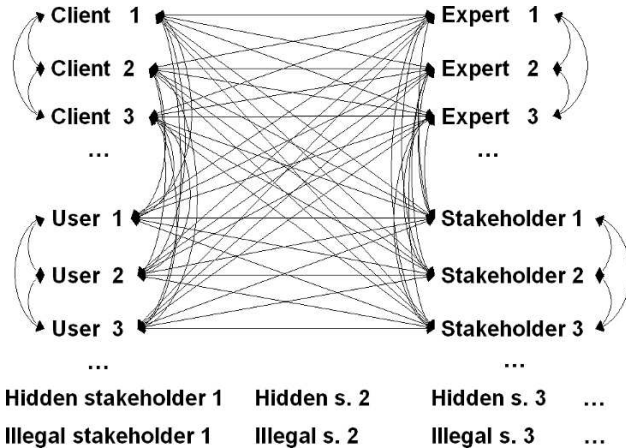
role explicitly and clearly defined as “the speculators”). In actual fact, many of them are “occult” and thus many of the relevant relations in the process of planning bypass the explicit and transparent “communicative arena”.

Scheme 7.



Finally, a possibility that frequently exists is that besides the “occult interests” (take as an example the classical Francesco Rossi movie *Le Mani sulla Città*, literally “*Hands on the City*”, where the web of relations and collusions appears in all its dramatic complexity), there are also illegal or criminal interests: “actors” who have indeed exercised an important role in many phases of many histories of urban development.

Scheme 8.



These long premises show how a planning process is actually extremely complex and highlight the relevance of the role of communication, negotiation and compromise.

2 What Is a Good Model?

It is beyond the scope of this article to discuss in detail all the questions that would need to be discussed, but we would like to underline – even if we limit ourselves to some of the previously mentioned aspects only, and in particular to the relationships between an expert (the modeller), other experts (the planners), all clients and all users – how necessary it is to be aware of such a complex web of relations, as we need to embed and enact within it the competencies and knowledge of a good modeller. With this awareness, we should extract the characteristics of a good model.

Dealing generally with the concept of *bounded rationality* proposed by Simon, and the entire discussion on complexity, the substantial outcome of which was to propose models based on a bottom-up rather than on a top-down paradigm, we believe the concept of *ecological rationality* should also be included (see Gigerenzer et al. 1999), suggesting that a heuristic element should be considered rational if it adapts well to the structure of its environment; what is needed then is a “tool-box” with “fast-and-frugal” tools which work well and rapidly.

Some years ago (Cecchini 1999), we proposed the following characteristics of a good model:

1. it should not be a black-box; it is essential that those who use it for planning as well as those at whom the plan is directed, understand how it works and why;
2. a model should predict and take into account actions and reactions of social actors as well as their interests, conscious or not, disclosed or not, rational or not;
3. a model should be such as to enable the assessment of as many alternatives as possible, as well as the understanding of them and the differences between them;
4. a model should be compatible with other models, even if differing in formulation and techniques used;
5. a model should be parsimonious, should not require an excessive number of variables, an excessive amount of data or excessive computational power;
6. a model should be flexible for different situations and contexts, and should allow fuelling, processing and handling with what is at hand;
7. a model should be fast to build, at least with respect to the schedules of the project the model is built for;
8. a model should be re-usable and anyway should never be a *hapax legomenon*.

In other words: we need models that are *modular* and *friendly*, i.e. all the planning protagonists should be able to use and understand them (avoiding the use of black-boxes, unless strictly necessary); models should be *flexible*, i.e. adaptable to different circumstances and to different phases of the planning process; *multi-level*, i.e. usable and useful in coping with different levels of complexity and interoperability; *low-cost*, i.e. of common use, usable when they are required and costing a small fraction of the total planning process costs; furthermore, they should be designed in such a way as to permit them to be *linked to communication*

tools of all kinds (or more precisely, of the “right” kind for each particular objective and for each particular group of actors); these and such models should serve:

- a. *to decide*, making the process transparent, motivated and responsible;
- b. *to negotiate*, allowing confrontation and comparison of different points of view and a possible composition of interests;
- c. *for consensus building*, in favour of a consensus that is informed and aware of the decisions taken by the public actors;
- d. *to evaluate*, verifying the results of the decisions and the *in itinere* correction of ineffective and contradictory decisions.

These models should reach their goals. And they must achieve these goals even when they are just a “part” of more sophisticated, “hard” ones used by experts, if and when necessary.

Finally, the objective of the analyst/modeller is to provide the actors involved in the planning process (and we have seen how numerous they are) with the ability to read, understand, represent, predict the different systems and domains of their actions.

It is not a simple task to have good models, even for, let us call them, epistemological reasons. In fact, the difficulty in dealing with the complexity of urban systems and the related difficulty in analysing and forecasting is twofold: the first lies in the complexity of the system itself, and the second is due to the actions of actors, which are “acts of freedom”.

3 Why Planning Is (Still) a Necessity and Why It Needs Models

We are convinced that planning *stricto sensu* (meaning by definition the production of “plans”) is an absolute necessity, today as always, nowadays more than ever.

We will not spend many but just a few words on justifying this position, which appears to us somewhat a minority one, but still very convincing.

The profession of planner has been changing over time, but it has experienced a particularly dramatic and rapid evolution during the last 20 years: a mutation that has substantially disrupted the “good principles” of classical urban planning, maintained solid for more than a century, even though with notable variations (see Hall 1988; Alexander 1992).

Among the causes of this change, an important role was surely played by technological innovations (Castells 1996, 2000), but primarily there has been a radical change in the social role and principles of the profession, determined by global affirmation of the so-called “unique thought” (*pensée unique*) which, according to

one of its greatest inspirers and creators, holds that “there is no such thing as society”².

If there is no such thing as society, the role of the planner disappears and loses its character. This idea is at the origin of a vision maintaining that spatial and territorial organisation is not the result of collective choices and an expression of the “general will” (society, in fact, with the power relations between different interests defined case by case), but rather the result of local negotiations between, on one side, the financial capital and the landowners (the owners of the “*jus aedificandi*”, elected and erected to the status of an inviolable right), and on the other, those who only can beg for a few concessions and impose few limitations on the full and uncontrolled exercise of such right. The crisis of cities and their loss of identity has its origin and causes above all in this. Planners, urban planners, could have done little to stand up against such an “aggression”, but the sad fact is that material disarmament was accompanied by intellectual (self-)disarmament.

Notwithstanding this, planning must be carried out, surely in new forms, too, less top-down, more participated, more flexible, more reversible, more environmentally attentive.

And indeed – we wish to say – planning can be carried out, for the idea seems to have been retrieved that a “common good” exists, and that it is possible - for experts, clients, users, organised or unorganised interests (explicit or hidden, legitimate or illegitimate, moral or immoral) (Simon 1981)³, with compromises, compensations, penalizations and repressions - to be a part of the government process, with a defined role.

And all this – we repeat – needs good models and models of a new kind: participated, flexible and reversible planning needs adequate models that are powerful, user-friendly, low-cost, modular, communicable and transparent.

That is why our profession as “creators of models” cannot avoid specifying (models) *for* planning, and this – much more than in other cases – forces us to work on making the links much stronger between techniques and objectives, between tools and their users.

Technological innovation, together with the development of a new epistemological vision of model building (Allen 1997), makes this objective possible.

We intended to spend more words than usual on theoretical speculation, which we do believe to be of cardinal importance in order to not make the presentation of models of a new conception and a new conception of models *for planning* futile.

But theoretical reflection is of little use without examples and good practice: this article will attempt to present one in particular, while just briefly illustrating a few others.

² “There is no such thing as society: there are individual men and women, and there are families.” Margaret Thatcher British Prime Minister, quoted by Bauman Z (1999); originally: interview on September 23 1987 for *Woman's Own*, published October 31 1987.

³ See also the Code of Ethics and Professional Conduct of the Planner (by the American Institute of Certified Planners): “A planner’s primary obligation is to serve the public interest. While the definition of the public interest is formulated through continuous debate, a planner owes allegiance to a conscientiously attained concept of the public interest”.

4 A Tool-Box for Planning

A good approach is the one proposed by Godet (1986), suggesting “seven key ideas for action and anti-fatality”:

- clarify the present action in the light of future horizons
- explore uncertain and multiple futures
- adopt a global and systemic vision
- take into account qualitative factors and actors' strategies
- always remember that information and predictions are never neutral
- choose pluralism and complementarity of approaches
- enquire into prejudices.

This is precisely the spirit of our tool-box, a set of tools for action for a good planner, a tool-box that may be used by experts, clients and users: not a black-box, but a process for understanding, learning and sharing.

We will here briefly describe a few experiences of ours in recent years. Afterwards we will concentrate on a tool we like a lot, that works better than others, a framework and piece of software for the construction of scenarios we have called *The Time Machine*.

4.1 Games and Gaming-Simulation

Games and gaming-simulations have proved to be an excellent learning and communication instrument: they can be used as a vector of knowledge based on real and complex systems, and also as a vector of skills and values. Games may even be used as a context for negotiation and as a place for revealing personal systems of preferences.

A game is “a pleasant activity in itself, it does not require any further objective beyond that”, writes Kant in *The Critique of Judgement*. It is a fully convincing argument. But that does not suffice, for what if the definition gives the necessary but not sufficient criteria? Each of us can recall activities “pleasant in themselves” different from games: numerous and different with respect to ideologies, experiences, ideas of the world, individual fantasies and perversions. We need to extend the definition: a “game” is a pleasant activity that has no other aim than itself; it is subjected to rules, but chosen freely; it takes place in a simulated world, and proposes victory as its main objective.

Clearly, therefore, educational games, communication games, participation games in themselves do not exist. A game is a game is a game?, and as such must maintain the peculiar trait of “autonomy of objectives” that a game must have.

Another issue is whether or not games *can contribute* to educating, to communicating or to stimulating participation. The answer here is affirmative: one can think for example of using a game to communicate, but the game used has to be a real game, and its first objective is to create a world, an environment, a behaviour, developing the taste for pleasure and, with it and within it, curiosity, courage,

fighting spirit, co-operation, determination, letting even cruelty appear, if necessary.

Only then, will the game operate in this indirect, unconscious but efficacious manner, as a game.

We have designed and developed several games in the past dealing with planning and environmental issues.

Let us start with a couple of simple games used for two exhibitions on urban sprawl (see Font et al. 2004; Indovina 2005). If one looks at them, their simplicity is easily noted; if one plays them, their “usefulness” is rapidly appreciated. The first game “Urban Animal” “only” serves the purpose of bringing citizens (the “users” in our scheme) closer to the themes of urban life: the resulting profile of the players (“what urban animal are you?”) might constitute an opportunity to reflect upon the multidimensionality of the contemporary city.

The “Future cities” game has a twofold function: on the one hand to present a set of urban projects proposed for a specific urban area (Barcelona and Bologna in our cases), and on the other to propose to users the idea of interdependency between various projects in making a desirable future scenario possible. The player has a defined amount of resources available, and after picking his/her “desired” future scenario, he/she needs to put together the actions (“implementing” projects each having a cost) he/she believes useful for achieving that specific scenario; a simple simulation model says if the operation was successful or not.

This set of games is a generalised *framework* that can easily be adapted and modified simply by substituting the files with questions, profiles, projects, scenarios, multimedia content⁴.

Another interesting experiment was the Como Mobility **GioCoMo** game. It is a simulation game for participating in metropolitan mobility in the city of Como which we named GioCoMo. The Municipality of Como commissioned the Milan University/Polytechnic to research the feasibility and possible alternatives of a tram-based transport system (see Colorni et al. 2003), for which our Laboratory had the task of creating, designing and developing a system for interactive communication oriented towards public consultation and the participation of citizens. The Web-based game GioCoMo was a part of this system (Blecic and Cecchini 2002)⁵. GioCoMo is a game that creates a relationship between clients, users and experts.

Another interesting example is the “Piazza Catalana” project, originally designed to promote the Catalan language in Alghero (Sardinia). The project’s infrastructure is grounded on an Internet-based 3D multi-user environment. Three environments (virtual worlds) have been designed with this system: a realistic 3D reconstruction of the present-day historic town core of Alghero, a historic reconstruction of Alghero at the beginning of the 16th century, and a future environment with no physical resemblance to the current town.

⁴ To see the game logon at www.lamp.sigis.net/games

⁵ For the Italian version see Blecic and Cecchini (2003a); for more information about GioCoMo logon at www.sigis.net/participation



Fig. 1. “Piazza catalana”: 3D reconstruction of present-day Alghero

This infrastructure could easily become a virtual space for information on, communication of and participation in planning in Alghero.

But probably the most interesting example we have developed recently is the game **SOS**. It is a multi-user Web-based game designed with the intent to communicate and transmit at least a part of the intrinsic complexity related to problems of government and environmental policies.

In recent years, mainly due to the spreading of Internet, many “video-games” have appeared – animated quizzes, pseudo-arcades or other genres – with the explicit objective of environmental education. Even if multi-medially rich, many such games offer an over-simplified and naïve idea of the problems and issues involved. This flaw is particularly visible in the way individual “environmentally-friendly” behaviours are encouraged and sponsored, without transmitting the expensiveness, and all the psychological, economic and social difficulties their adoption may imply. But above all, without communicating that a radically different government and use of environmental resources would require a remarkable political effort and systemic intervention, in comparison with which the issue of “good” individual behaviour has no more than a homeopathic function.

The game SOS is thus an attempt to provide “environmental education” by presenting the complexity and modalities of interweaving individual actions with public policies at a national, international and global level.

In the game, a player takes the role of a “representative” citizen of a nation, and has the chance to undertake actions and decisions, both those regarding his/her individual behaviour, and those that bring about interaction and negotiation of policies with other players, influencing the social, economic and political order of their nation-states.

Before choosing the nation-state to live in, players must do a test which helps determine their personal profile. This profile, measuring the player’s propensity to save or consume (parsimony vs. dissipation) their level of local or global identity (local vs. cosmopolitan identity), as well as their political orientations and opinions, is used in the game, among other things, to calculate players’ pay-offs after their action.



Fig. 2. General game interface

The simulated imaginary world is quite diversified, and nation-states can be rich or poor, a democracy or dictatorship, tolerant or fundamentalist, with different economic structures, different ways of using environmental resources and different ways of distributing richness and offering opportunities to their population. Even if it is only an invented world, this is pretty much the picture of our real world today.

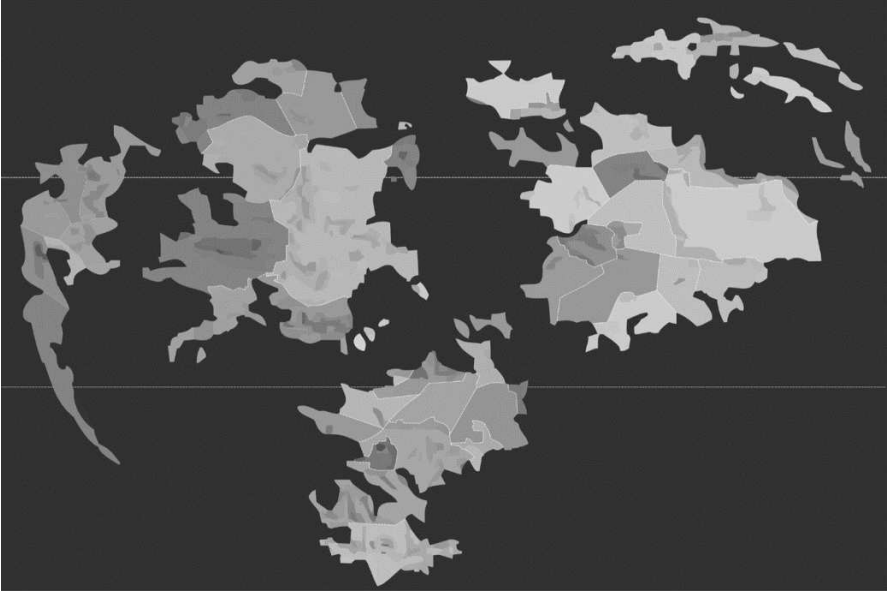


Fig. 3. The “world” map

SOS is a turn-based game, where each turn, representing a year in the life of the simulated world, lasts one day for a total of 25 turns. During each turn, players take decisions and choose options about their individual behaviour and decisions, express their opinion about general economic, political and social issues and can jointly contribute in financing and activating public policies. The game offers tools for interaction among players: discussion forums, chat and internal messaging.

The underlying model tries to simulate all the feedback dynamics on an individual level⁶ and on a societal level⁷. For these reasons, the relevant parameters of a nation (and part of the player’s performance) are not only environment-centred, measuring “environmental quality”, but also include the GNP *per capita* and its distribution, the state of human and political rights and the structure of the economy.

⁶ For example, the decision to use public rather than private transport may create different personal satisfaction, depending on the player’s personal profile and on the state of the nation.

⁷ For example, a transition to an integrally “clean” economy has its economic, social and political impacts.

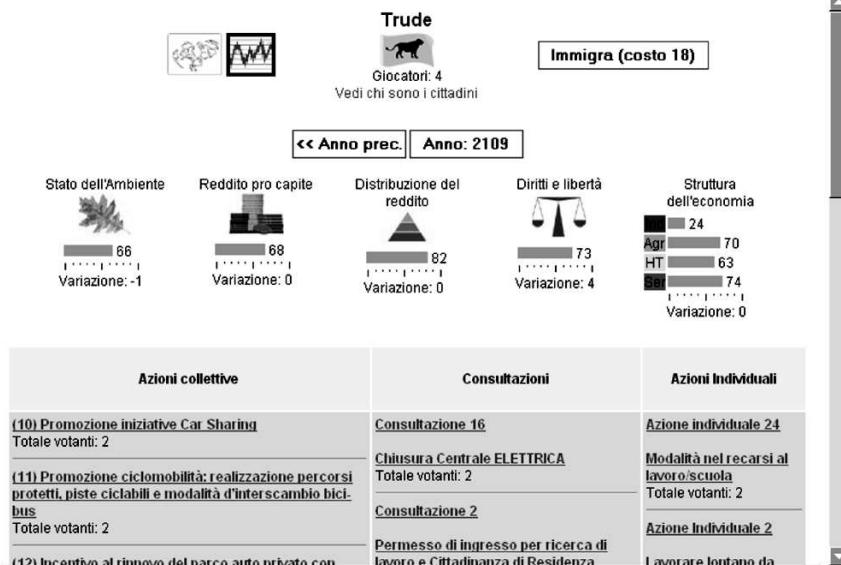


Fig. 4. Nation status

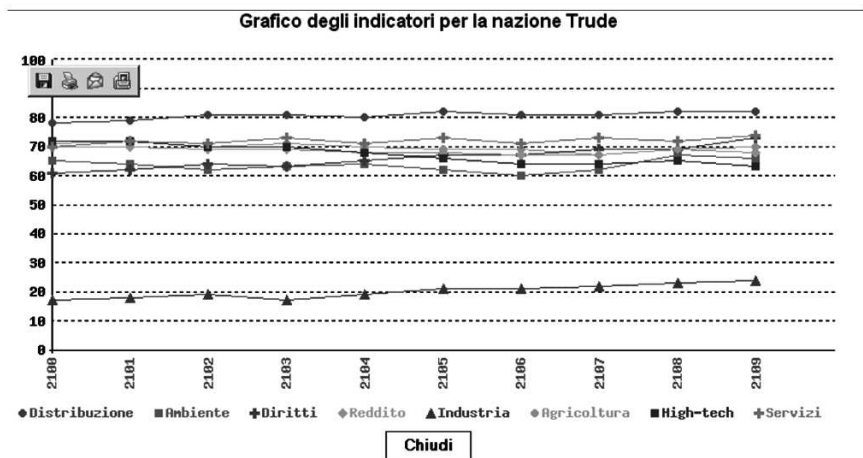


Fig. 5. Evolution of nation parameters

To survive the development of (or in the development, or against the development, or notwithstanding the development, or with the development) and to construct a sustainable society is a goal that imposes understanding and management of the complexity, both intrinsic to systems with many objects interacting in a non-linear manner, and due to the freedom of action of individuals.

We are convinced that this game is a useful contribution to the objective of designing and developing tools that do not tend to simplify reality, but rather try to entirely employ its complex nature, on a behaviour and values level, too.

4.2 Global Interactive Maps for Learning (MaGIA)

Getting closer to the central topic of our contribution, we will just mention here two other tools that are in the development and testing phase and which – we believe – adopt the same logic and which we would like to make interoperable with all the others. One is a sophisticated on-line multi-user system for collective construction of concept maps that we have called Global Interactive Maps for Learning (*MaGIA - Mappa Interattiva Globale di Apprendimento*). The second is an environment for cellular automata simulation named *Cellular Automata General Environment (CAGE)* which permits us to build models and carry out simulation of the effects of planning decisions.

MaGIA is a tool for interaction between clients, experts, users and interest groups, while *CAGE* is more oriented towards the relationship between clients and experts.

MaGIA is Internet-based software for *collective* construction of concept maps. The grounding hypothesis of the software is the idea that it is possible to represent collective “knowledge” on a particular topic through a network where concepts are represented as nodes and where the links between them are represented as arcs.

It is a tool for bottom-up construction of concept maps, where every user can intervene and contribute to the construction of the map by adding concepts and links between them.

For each concept and for each link it is then possible to specify definitions, add documentation, bibliographical and web-biographical references.

The system also includes features for communication among participants, in a synchronous (chat, instant messaging) and asynchronous (discussion forum) way.

In order to make a “selection” within concept maps, potentially extremely rich but also full of “white-noise” in cases with a large number of participants, each participant has the possibility to express his/her vote (of preference, adequacy, correctness, ...) for every single element of a concept map, including nodes, arcs (enforcing or loosening in this way the relations between nodes), definitions and other elements other users have inserted in the system. With this information, it is possible to activate different types of filters and apply a “natural selection” between all the elements of a map and extract a “dominant” map.

The software offers a series of tools for the analysis and graphical representation of knowledge maps. Hence, it is possible to visualise the map as a bi-dimensional or tri-dimensional network that can be used as a means of navigation from concept to concept.

Furthermore, several types of filters can be activated on maps thus visualised, for example:

- visualise the connections between concepts where the “strength” of connection exceeds a particular value defined by the user;
- activate – starting from a “concept of aggregation” series defined by the user – a cluster analysis;
- visualise “hierarchies” between concepts, where the importance of a concept depends on the number and strength of connections with other concepts;
- search for particular “geometry” that might be of particular interest.

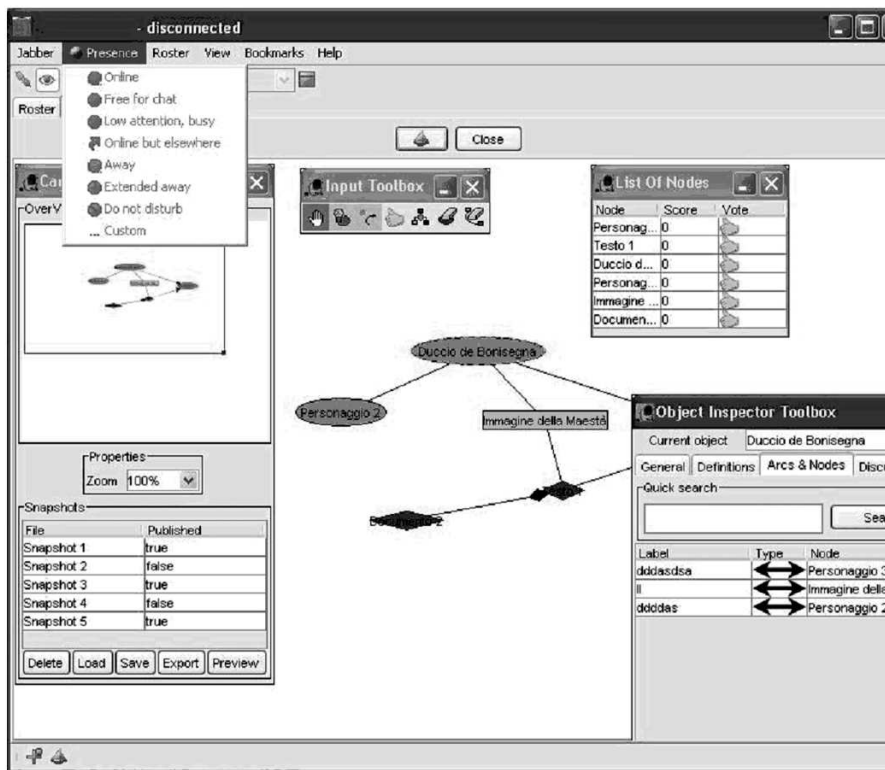


Fig. 6. MaGIA general interface

The use of the system can – in principle – cover a wide range of purposes, from a careful theoretical discussion to support for collective *brain-storming*.

Furthermore, of particular interest is the collective and “horizontal” nature of constructs obtained with this system. Thought up and designed as intrinsically democratic and egalitarian, it can reveal its usefulness as an instrument of activation of citizens’ participation in public decision-making.

In contrast with many “solutions” for participation where a greater emphasis is put on “vertical” relationships (traditional decision-makers → target groups), our

system is designed around horizontal interaction which can give origin to more qualified and reasoned thoughts, proposals and hypothesis.

On the other hand, as the name itself suggests, the software can be considered a formidable tool for collective learning and knowledge transmission. The learning process does not advance only via consultation of a hyper-medial encyclopaedia, but also through active participation, through interaction between those who offer to be “depositors” of local knowledge, no matter how detailed or partial.

For a more detailed presentation of *MaGIA* see Blecic and Cecchini (2003b)⁸.

4.3 Cellular Automata Generalised Environment (CAGE)

The interdisciplinary nature of CA-based modelling is an important challenge, resulting in an increasing demand for dedicated Problem Solving Environments (PSEs) suitable for supporting educational, research and industrial activities. The main objective of these software environments is to relieve the researcher from low-level operational and programming activities, allowing him/her to concentrate his/her efforts on the modelling itself. Furthermore, the visualisation and data analysis features usually available in PSEs make the CA modelling-calibration-validation cycle much faster and effective. As the main consequence of all this, PSEs are of particular use in situations where an integration of different interdisciplinary knowledge is required, which in CA modelling of real systems typically corresponds to the involvement of researchers from various scientific areas, and not just the computer scientist.

As pointed out by Worsch (1999), there exist a number of software packages for CA-based simulations. Some of these implemented variations and generalisations of the classical definition of CA, have been proved to be useful for modelling real phenomena. On the other hand, many studies related to the application of CAs to the simulation of dynamic processes in geography (Batty and Xie 1994; Batty 1997; Cecchini 1996), have stressed the need for a stronger relaxation of classical CA assumptions with respect to those usually implemented in available CA environments.

First of all, even if the great majority of CA applications successfully adopt a strict local neighbourhood, spatial interactions in geographical phenomena can often take place over greater distances. Hence, the CA model we present herewith is partially based on the notion of *proximal* space, deriving from research in ‘cellular geography’ (Tobler 1979) which set the basis for the so-called *geo-algebra* approach proposed by Takeyama and Coucleis (1997). In this later work, relaxation of some of the main assumptions of classical CA theory has been suggested, namely, spatial regularity of the lattice and homogeneity of the cells’ neighbourhoods. Thus, in *geo-algebra* every cell has a different neighbourhood defined by relations of *nearness* between spatial entities (i.e. cells), where *nearness* can mean both, topological relations or generic “behavioural” (e.g. functional) influence. Following the approach described, in the present work we shall define neighbour-

⁸ For more on *MaGIA* logon at www.lamp.sigis.net/participation.

hood as a generic set of cells the state of which can influence, in some way, the variation of the current cell's state. Another relaxation we have adopted regards the temporal stationariness of neighbourhoods which, contrary to classical CAs, can change with time.

In comparison with most CA models, we have introduced a further variation inspired by the observation that the simulation of complex systems is often founded on the interaction between quite different sub-systems. Such modelling might require different sub-systems' dynamics to be based on knowledge from different disciplines, and can require different spatial and temporal granularity. Thus, likewise in Bandini and Mauri (1997), it appears useful and effective to organise the 2D-physical domain in more distinct *layers*. Cells can have both a *horizontal neighbourhood*, a subset of cells belonging to the same layer, as well as several *vertical neighbourhoods* with cells belonging to other layers.

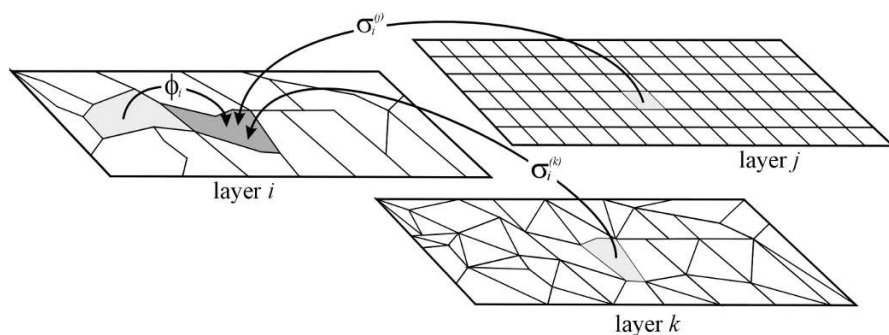


Fig. 7. The multi-layer structure of the AC model implemented in CAGE

Moreover, with respect to the classical CA formulation, the proposed model includes many of the well-known generalisations and modifications (e.g. see Di Gregorio and Serra 1999), such as parametric dependence of transition rules, which simplify global steering of CA (i.e. global-to-local influence), and cell state decomposition in sub-states.

Another relevant issue is the integration and interoperability of available geographical data infrastructures with CA environments. The integration of CA engines into existing GIS systems has been proposed (e.g. loose coupling based on the Remote Procedure Call paradigm or other proprietary protocols) (Clarke and Gaydos 1998; Wu 1999; Wagner 1997). However, for the CA reference model proposed herewith, loose coupling with proprietary GIS is hard to obtain both for the required flexibility, especially in the modelling phase, and due to low computational efficiency.

These considerations have induced us to develop a modelling environment based on a widely generalised CA formulation. The environment, described with greater details in the paper, has proved to be particularly suitable for the modelling and simulation of spatial urban, territorial and environmental phenomena.

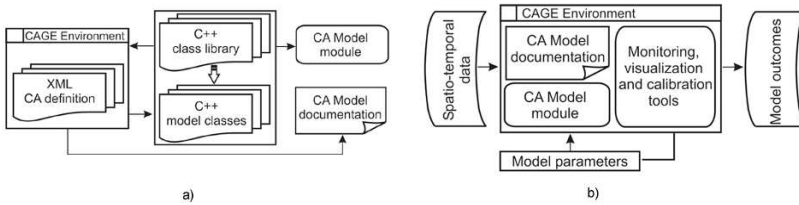


Fig. 8. a) Modelling and b) simulation of a cycle in CAGE

Currently, after testing the prototype version (see Fig. 9), we are now developing an “operative” and stable version of the software.

For a more detailed presentation of CAGE see Blečić et al. (2004a); for a case study see Blečić et al. (2004b).

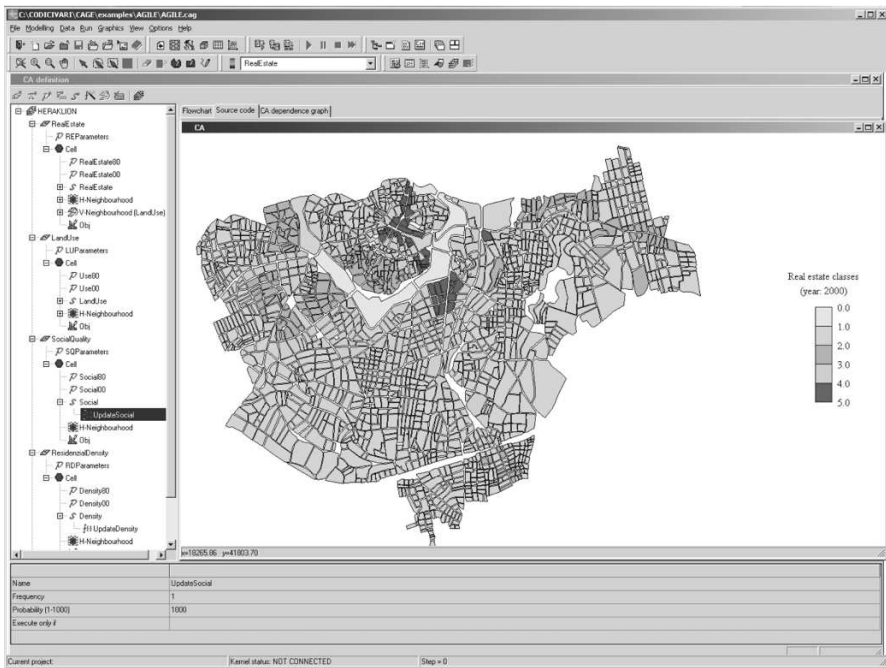


Fig. 9. A screen capture of a CAGE simulation

5 The Time Machine

Let us finally have a closer look at the methodological framework and respective software tool that we have developed and named *The Time Machine*. It produces

now-for-then future scenarios, meaning that time is not assumed as a continuous interval, but is stylised as a predefined time leap. The scenarios produced are based on the interaction of four types of entities that build every *Time Machine* model.

First of all, there are the so-called *Events*. They are the essential ingredient of every *Time Machine* model, and should be considered the atomic events thought to be possible within the given time frame. Their definition must clearly include the estimated probability of occurrence within that time interval.

Another type of entity we have called *Variables*. These are true quantitative variables, and are thus expressed in the appropriate unit of measurement.

Furthermore, there are two other types of entity. The so-called *Exogenous Events*, which are similar to *events* inasmuch that they are characterised by the probability of occurrence, but logically represent influential events that are external (exogenous) to the modelled system. The consequence of this external status is that, as we shall see in greater detail below, they influence other entities of the model, while not being influenced in their turn.

Finally, the fourth and last type of entity is the so-called *Actions*, which represent possible actions the decision-maker can put into practice or implement. Every *action* is characterised by an "effort" value, measuring the "cost" of the action in its wider sense (economic, social, electoral, normative, consensual, etc.) In this frame, we shall call *strategy* a set of actions implemented jointly.

All the entities are mutually interdependent and the "influences", expressed through a set of functions, define how the "occurrence" of every single entity during the simulation modifies the probability or the value of other entities related to it. We will see later the exact meaning of the term "occurrence" for each of the four types of entity. The general overview of interactions among entities is illustrated in Fig. 10.

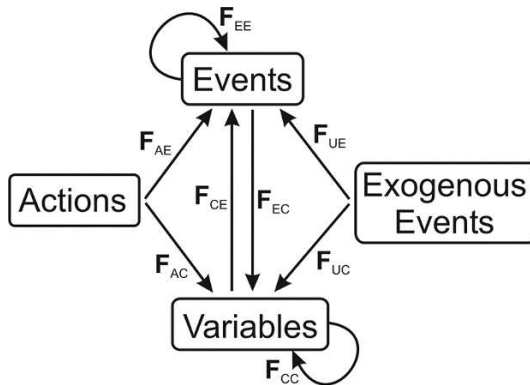


Fig. 10. The general overview of interactions between entities of *The Time Machine Model*

As can easily be seen, there are a number of issues related to the building of an effective *Time Machine* model: the definition of entities, the time "interval", probabilities, efforts, influences.

This process of definition of a model can be done directly by "experts", or can be the fruit of a collective discussion between experts, clients, users and interest groups; in the latter case, the model-building process itself is useful for the common understanding of problems and issues, their conceptualisation as well as the exploration of the fields of possible solutions.

5.1 The *Time Machine* Model: a "Light" Formal Definition

In formal terms, we can describe the basic model as follows. The entities the system is made of are collected in a quadruplet:

$$\Sigma = \langle \mathbf{e}, \mathbf{u}, \mathbf{c}, \mathbf{a} \rangle$$

where \mathbf{e} is the vector of *events*, \mathbf{u} the vector of *exogenous events*, \mathbf{c} the vector of *variables* and \mathbf{a} the vector of *actions*. In particular:

- each component e_i in the vector \mathbf{e} is an *event* defined by a probability of occurrence $\hat{\Pi}(e_i)$ estimated under the assumption of complete "isolation" (i.e. with no influence from other entities in the model);
- each component u_i in \mathbf{u} is an *exogenous event* characterised by a probability of occurrence $\hat{\Pi}(u_i)$, which by definition of exogenous events cannot be influenced during the simulation, and is therefore a constant value;
- the generic component c_i in the vector \mathbf{c} is a *variable*, defined as a scalar variable defined in an interval $I_c = [c_i^{\min}, c_i^{\max}]$ and characterised by an *initial value* \bar{c}_i ;
- the generic component a_i in the vector \mathbf{a} is a scalar variable representing the *effort* of an *action*, defined in an interval $I_a = [a_i^{\min}, a_i^{\max}]$;

The interactions among entities are defined by a set of impact factor matrices and by specific laws of variation. There are two groups of matrices, deriving from the fact that in one case during the simulation the *probabilities of events* are influenced and changed, while in the other case it is the *value of variables* that is being modified. In particular, we have:

- the impact factor matrices \mathbf{F}_{UE} , \mathbf{F}_{EE} , \mathbf{F}_{CE} and \mathbf{F}_{AE} the generic element of which $f_{ij} \in [-f_E, f_E]$ determines the variation of the probability of the *event* e_j , due to the occurrence respectively for each matrix of the *exogenous event* u_i , of the *event* e_i , of the variation of the *variable* c_i and of the implementation of the *action* a_i . The exact mechanism of variation is explained further below.

- the impact factor matrices \mathbf{F}_{UC} , \mathbf{F}_{EC} , \mathbf{F}_{CC} and \mathbf{F}_{AE} the generic element of which $f_{ij} \in [-f_C, f_C]$ determines the variation of the value of the variable c_j due to the occurrence respectively for each matrix of the *exogenous event* u_i , of the *event* e_i , of the variation of the *variable* c_i and of the activation of the *action* a_i .

For the effects on the probabilities of factors collected in matrices \mathbf{F}_{UE} , \mathbf{F}_{EE} and \mathbf{F}_{AE} , the assumption is that the impact factor f increases or decreases the probability of *events* in an indirect way. This is obtained through a relation expressing *rigidity* to variation. A possible relation for this could be the following:

$$\Pi^*(e_i) = \begin{cases} \Pi(e_i) + \frac{1 - \Pi(e_i)}{f_{MAX}} \times f, & f \geq 0 \\ \Pi(e_i) \times \left(1 + \frac{f}{f_{MAX}}\right), & f < 0 \end{cases}$$

where $\Pi^*(e_i)$ is the *updated* probability of the event.

The previous expression has a specific and desired feature: the resistance to change grows as the value of the probability draws near its limits (and thus the effect of the impact depends on the state the system is currently in). It is easily shown that the factor $f/f_{max} > 0$ would represent the *increase* of probability of the event, had its prior-to-impact-probability been 0, while $f/f_{max} < 0$ would decrease the probability of the event, had that probability been equal to 1. Such formalisation helps to define the impact factor in a range with only few values (e.g. in the interval [-5, +5]).

Every element in the matrix \mathbf{F}_{CE} (which determines the effect on probabilities of events of a variation Δc_i of a generic *variable* value) is used in a relation analogous to the previous, except that it includes prior multiplication by the factor $\Delta c_i / (c_i^{\max} - c_i^{\min})$.

As far as the effects on the *variables* are concerned, it is assumed that the impact factor f acts on the transformed quantity $q_i = (c_i - c_i^{\min}) / (c_i^{\max} - c_i^{\min})$ via a relation analogous to that of probabilities (note that $q_i^* \in [0,1]$). Once the updated value q_i^* has been calculated, the corresponding value of the variable c_i^* is calculated with the inverse formula:

$$c^* = (q^* \times c^{\max} + c^{\min}) / (1 + q^*)$$

The impact factors in the matrices \mathbf{F}_{UC} , \mathbf{F}_{EC} and \mathbf{F}_{AC} are used directly, while those in the matrix \mathbf{F}_{CC} are preventively multiplied by the factor $\Delta c_i / (c_i^{\max} - c_i^{\min})$, being Δc_i the variation of the variable in question.

The simulation procedure uses a Monte Carlo-like algorithm (as suggested originally in Gordon and Hayward 1968), which may be repeated (iterated) many times. The basic version of the execution procedure for a single iteration is described in Algorithm 1.

At the end of each iteration of this simulation procedure, all events will have a value 0 (not occurred) or 1 (occurred), while *variables* will have reached a final value.

By repeating the iteration many times and by calculating the average values, we can obtain the so-called *experimental probabilities* of *events* and the *dynamic average value* of *variables*. For the scenario analysis the possibility is of particular interest of comparing these values with the initial probabilities or values, and with the probabilities and values resulting from the sole "implementation" of *actions* (and thus considering only first-level effects of actions).

1. Put all the *actions* "implemented" by the user, all the *events*, all the *exogenous events* and all the *variables* in a single collection.
2. Extract randomly one entity from that collection:
3. If the extracted entity is an *action*, calculate the variation of probabilities of *events* and the variation of values of *variables*, due to the effects of the *action* based on the relation matrices \mathbf{F}_{AE} (actions \rightarrow events) and \mathbf{F}_{AC} (actions \rightarrow variables)
4. If the extracted entity is an *exogenous event*, then test if it occurs with regard to its probability ("throw the dice"):
 - If the *exogenous event* occurs, then calculate the variation of probabilities of *events* and the variation of values of *variables*, based on the relation matrices \mathbf{F}_{UE} (exogenous events \rightarrow events) and \mathbf{F}_{UC} (exogenous event \rightarrow variables);
 - If the *exogenous event* does not occur, do nothing;
5. If the extracted entity is an event, then test if it occurs with regard to its probability ("throw the dice"):
 - If the event occurs, then calculate the variation of probabilities of events and the variation of values of variables based on the relation matrices \mathbf{F}_{EE} (events \rightarrow events) and \mathbf{F}_{EC} (events \rightarrow variables);
 - If the event does not occur, do nothing;
6. If the extracted entity is a variable, then calculate the variation of probabilities of events and the variation of values of variables based on the relation matrices \mathbf{F}_{CE} (variables \rightarrow events) and \mathbf{F}_{CC} (variables \rightarrow variables).
7. Repeat phases 2 to 6 until all entities have been extracted from the collection

Algorithm 1. The inference procedure implemented in *The Time Machine*

More formally, considering a system with n_e *events* and n_c *variables*, the execution of m iterations of the simulation procedure would provide:

- A matrix \mathbf{K} , of the dimension $m \times n_e$, the generic element of which k_{ij} equals 1 if the event e_j occurred during the i -th iteration, otherwise k_{ij} equals 0;
- A matrix \mathbf{S} of the dimension $m \times n_c$, the generic element of which s_{ij} equals the value the *variable* c_j has assumed at the end of the i -th iteration.

The experimental probability of every *event* e_i can be estimated as the frequency of occurrence during the m iterations:

$$P(e_i) = \frac{1}{m} \sum_{j=1}^m k_{ji}$$

It must be noted that the matrix \mathbf{K} also permits us to estimate the probability of any *composite event* (defined as a subset of events), still using the calculus of the frequency of occurrence. For example the probability $P(e_i \wedge e_k)$ that two events e_i and e_k occur concurrently in a single iteration, or the probability $P(e_i \wedge \neg e_k)$ that the event e_i occurs while the event e_k does not occur.

The concept of *composite event* is very important since it is precisely what we in this framework will call *scenario*. In other words, any subset of events shall be called a *scenario* shall be called any subset of events, and the probability of such a scenario will be estimated as the probability of the *composite event* containing all the events from the subset.

The dynamic average value of a variable c_i is calculated as the average of the values at the end of an iteration based on the matrix \mathbf{S} :

$$\bar{c}_i = \frac{1}{m} \sum_{j=1}^m s_{ji}$$

As mentioned before, for the purpose of analysing what we could call the "system-effect", it is useful to compare the experimental probabilities of *events* and the dynamic average values of *variables* with those obtained through the execution of the Monte Carlo simulation with the sole implementation of *actions*.

5.2 The Inverse Problem: Searching for "Best" Strategies

The proposed model offers a heuristic opportunity of great practical value: if the events can be classified in terms of desirability (the simplest way is to classify them as positive, negative or neutral), *The Time Machine* software tool includes the possibility to activate the searching for strategies procedure that maximises the probabilities of positive scenarios while minimising those of negative scenarios. This search procedure is based on a *genetic algorithm approach*.

The problem can be formulated as follows. First of all, thanks to the cross-impact model outlined, we can assume that the joint probability P of a subset of events (i.e. the probability of a scenario) is implicitly expressed as the probabilistic function of the activated strategy. Remember that a *strategy* has been defined as a subset of *actions* implemented jointly, so it can formally be represented as a vector of bits $\mathbf{m} = (\mu_1, \mu_2, \dots, \mu_{na})$ where μ_i equals 1 when the i -th *action* has been activated, and equals 0 otherwise.

Subsequently, the vector of events can be projected onto two orthogonal subspaces thus producing two sub-vectors \mathbf{e}_g , collecting the n_g events considered positive, and \mathbf{e}_b , collecting the n_b events considered negative. Clearly, the two sub-vectors are composite events, in the terms exposed earlier.

The objective of the search procedure is to determine the strategy \mathbf{m} that maximises the probability of the composite event \mathbf{e}_g and minimises the probability of the composite event \mathbf{e}_b . The searching algorithm can furthermore include other auxiliary objectives such as the minimisation of overall implemented effort of *actions* or simply satisfy the requirement of maximum effort allowed. Therefore:

$$\begin{cases} \max_{\mathbf{m} \in \Omega} \delta_g(\mathbf{m}), & \delta_g(\mathbf{m}) = P(e_g) \\ \max_{\mathbf{m} \in \Omega} \delta_b(\mathbf{m}), & \delta_b(\mathbf{m}) = P(\neg e_b) \\ \max_{\mathbf{m} \in \Omega} \delta_\mu(\mathbf{m}), & \delta_\mu(\mathbf{m}) = \sum_{i=1}^{n_a} \mu_i a_i \end{cases}$$

with the constraint $\delta_\mu(\mathbf{m}) \leq \mu_{\max}$, where Ω is the space of the vector \mathbf{m} (which have 2^{n_a} elements) and μ_{\max} is the maximum allowed effort.

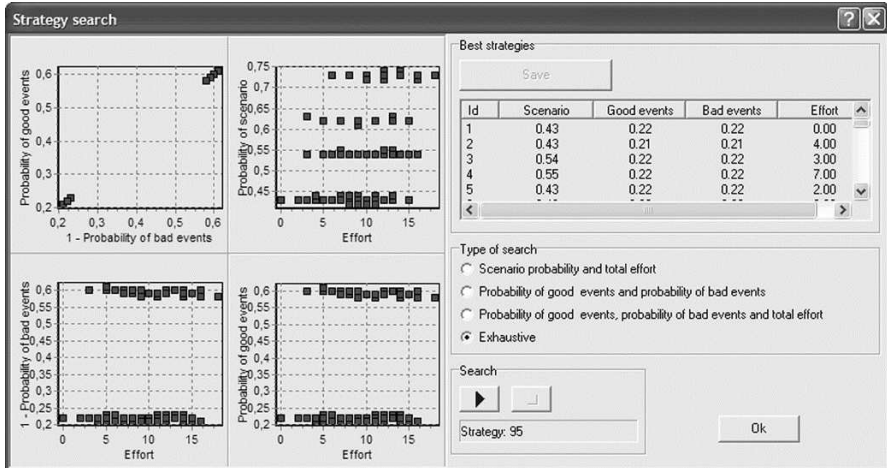


Fig. 11. Screen capture of the *Strategy Search* window

For a more detailed description of the genetic search algorithm implemented in *The Time Machine* software, see Blečić et al. (2005).

The objective functions of such a problem are not available in an explicit form: their value can be estimated only experimentally, by executing simulations. This makes it rather difficult to use some classical methods of optimisation, such as gradient-based ones, while suggesting the use of a heuristic technique grounded on "local" knowledge of the function, namely genetic algorithms.

In our case, a genetic algorithm is used to make an initially randomly initialised population evolve, whose generic element is a "chromosome" representing the above-mentioned vector \mathbf{m} . In particular, the problem is approached using the concepts of Pareto-optimality and dominance (Fonseca and Fleming 1995; Goldberg 1998; Srinivas and Deb 1995; Coello et al. 2002; Deb and Jain 2002). This avoids the need to make arbitrary assumptions about the relative importance of single objectives. Indeed, as widely discussed in the literature, a genetic search based on such a definition of optimality produces as a result not only one, but a set of non-dominated solutions, from which the decision-maker can choose the one considered most appropriate.

In a more sophisticated version of the model, the intensity of impact of each *action* on *events* and variables depends also on the effort invested, according to an *ad hoc* response function. In other words, *actions* are not simply implemented, but are a set of entities the user can invest more or less effort “points” in. In this version of genetic research, the analysis is oriented towards determining the effort values to allocate to each *action*.

5.3 Why Another Cross-Impact Analysis Software?

Ever since the first experiments done by Gordon and Hayward (1968), various software tools have been developed for cross-impact analysis-based simulation and scenario analysis, as well as for the search for strategies. Many of these tools implement different variations of the cross-impact analysis approach. As already remarked elsewhere (Krauth et al. 1998), in the fields we are dealing with an effective support tool should at least satisfy the following requirements:

- permit free modelling of the interdependencies between the entities the model is built of;
- include the possibility for modelling both the impact of external (exogenous) events, as well as effects of potential actions;
- offer the possibility to undertake both quantitative and qualitative analysis; have an effective and friendly user interface, simple to learn and use.

Unfortunately, many software environments that have been developed in the past do not include all these characteristics, and those that are more sophisticated are not freely accessible. All this has induced us to develop the *The Time Machine* software, which is a freeware tool developed in C++ language. The software runs locally under the Windows operative system, but was accurately developed to warrant painless execution under *Wine* (the widely used Open Source implementation of the Windows API for x86-based unixes, such as Linux, FreeBSD and Solaris) with little or no runtime overhead. Figure 12 is a screen capture from a *Time Machine* session.

The software was designed in such a way as to support an articulated but communicable description of the system's structure and of the interactions between entities in different decision-making scenarios. All that – we believe – makes this tool an effective and insightful support for debate, collective discussion and participatory processes.

The use of the software within such communicative contexts offers some fundamental advantages, such as user-friendliness, low-cost, modularity, high communicability and transparency, adequate also for participated planning contexts. Such characteristics are more and more of crucial importance for the success of evaluation, strategic planning and participation processes, intended in a wider sense as the occasion for creation of value on the local level and learning both on the local level as well as at the level of the institutional decision-makers.

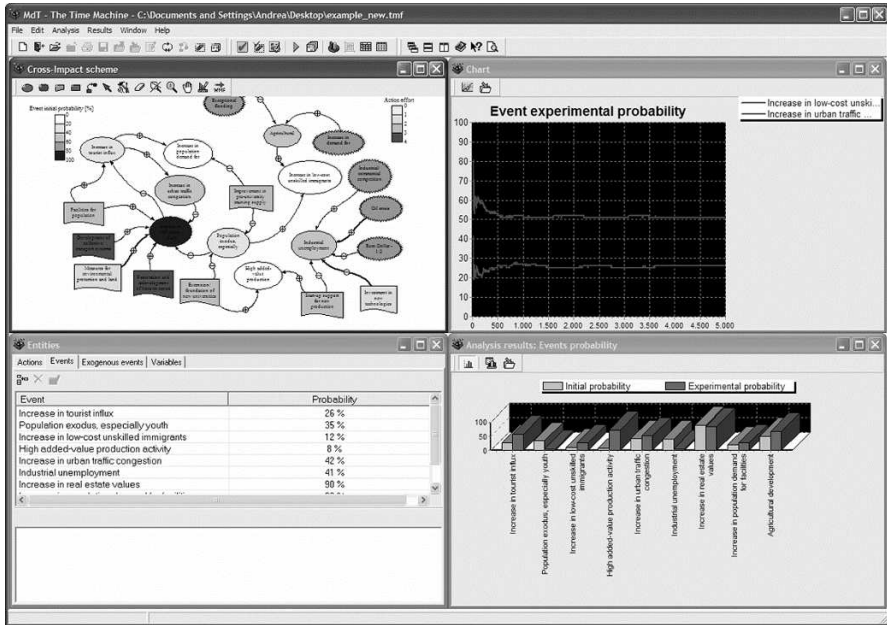


Fig. 12. Screen capture of *The Time Machine*

5.4 An Example

We will only briefly outline a model that will hopefully provide a concrete example of possible applications of the approach described.

The model is related to a research project on the evaluation of policies on tourism. For that purpose, we undertake a study of the typologies of tourist populations. Studies on profiles of tourist populations have been particularly intensive in recent years, as a necessary condition for the definition of actions and strategies for the purpose of sustaining or reducing the influx of particular categories of tourists; clearly, these “populations” are interconnected in complex manners, so for example frequently a policy intended to favour a certain type of tourism can also bring about the undesired effect of attracting other types, too.

The conceptual reference of our model is to define tourists using four distinct parameters (which in the simplified version can have only two modalities: “present” or “absent”). Their combinations thus give origin to 16 potential profiles. The four parameters are:

- *Loyalty*: is the tourist a habitual visitor to the place? (an example indicator might be related to the number of consecutive years of visits);
- *Duration*: does the visit last long?

- *Mobility*: does the tourist remain in a single place (e.g. a tourist village) or is he/she moving around the territory? (an example indicator might be related to the number of different places visited in a specific period of time);
- *Period*: does the tourist visit the place (or is willing to visit it) only during a specific period or does he/she visit it (or might be willing to visit it) during other periods of the year, too?

Clearly, the exact measurement of these criteria implies a rigorous definition of indicators and the availability of adequate data.

We have attributed evocative names to each of the 16 profiles deriving from the combination of the four characters. For example:

- L+D+M+P+ (or briefly LDMP) has been called *citoyen* (citizen) to indicate that – after all – a tourist that comes all the time, stays for long periods, moves around a lot and visits the place in different seasons is not much different from a resident;
- L+D+M+P- (briefly LDM) is a *constrained citoyen*, which while identifying himself/herself with the place, still cannot – probably for work – visit it during certain seasons of the year.

Clearly, the number of people belonging to a certain profile can change both for new tourists, as well as for tourists who switch from one profile to another; these variations may depend on external factors (e.g. the US Dollar /Euro exchange rate) or on internal factors, fundamentally due to the type and quality of tourist supply.

This conceptualisation can easily be transformed into a *Time Machine* model, where the *actions* are the available and possible policies, the *events* represent different aspects of territorial supply and quality, while the number of incoming tourists per each of the 16 profiles are defined as variables.

Conclusions

The Time Machine might look like a rather “light” approach to future scenario construction, and indeed a limited answer to the demand for forecasting techniques and tools. But it must not be underestimated due to the fact that it can be easily activated and modified by users, and that it provides an accurate and transparent analysis of the evolution of the system (as opposed to a “black box”); it should not be underestimated since it reminds us, with Godet, that “il serait dangereux de limiter la réflexion au scénario considéré comme le plus probable, car bien souvent ce dernier n'est en réalité que le moins improbable”⁹ (Godet 1984 p. 43)

⁹ “It would be dangerous to limit reflection to the scenario considered as most probable, because very often this latter is actually only the least improbable” (*our translation*).

It is thus a good example of that particular mix of empiricism, modesty, attention to the qualitative, a prudent *calcuemus*, attention for alternatives and counter-intuitive effects, that surely makes *The Time Machine* a potential candidate for the contemporary planner and designer's tool-box.

References

- Alarcòn LF, Ashley DB (1998) Project management decision making using cross-impact analysis. *Int. Journal of Project Management* **16**: 3: 145-152
- Alexander ER (1992) *Approaches to Planning: Introducing Current Planning Theories, Concepts and Issues*. Gordon and Breach, Philadelphia
- Allen PM (1997) *Cities and regions as self-organizing systems: models of complexity*. Gordon and Breach Science, Amsterdam
- American Institute of Certified Planners, Code of Ethics and Professional Conduct of Planner
- Bandini S, Mauri G (1999) Multilayered Cellular Automata. *Theoretical Computer Science* **217**: 99-113
- Bauman Z (1999) *In Search of Politics*. Polity, Oxford
- Batty M (1997) Urban systems as cellular automata. *Environment and Planning B* **24**: 159-164
- Batty M, Xie Y (1994) From cells to cities. *Environment and Planning B* **21**: 31-48
- Blecic I, Cecchini A (2002) GioCoMo. In: Percival F, Godfrey H, Laybourn P, Murray S (eds) *Interactive Learning through Gaming and Simulation. The International Simulation and Gaming Yearbook*, vol 11
- Blecic I, Cecchini A (2003a) GioCoMo: un gioco basato sul Web per la partecipazione pubblica. Proceedings of the INPUT 2003 conference, Alinea Editrice, Florence
- Blecic I, Cecchini A (2003b) MaGIA: mappe globali interattive di apprendimento. In: Maciocco G, Pittaluga P (eds) *Immagini spaziali e progetto del territorio*. FrancoAngeli, Milan
- Blecic I, Cecchini A, Trunfio GA (2004a) A Generalized Rapid Development Environment for Cellular Automata Based Simulations. In Sloat PMA, Chopard B, Hoekstra AG (eds.) *Lecture Notes in Computer Science, Cellular Automata: 6th International Conference on Cellular Automata for Research and Industry, ACRI 2004*, Amsterdam
- Blecic I, Cecchini A, Prastakos P, Trunfio GA, Verigos E (2004b) Modelling Urban Dynamics with Cellular Automata: A Model of the City of Heraklion. In: Toppen F, Prastakos P (eds) *AGILE 2004 Conference Proceedings*, Crete University Press, Heraklion
- Blecic I, Cecchini A, Trunfio GA (2005) A Decision Support Tool Coupling a Causal Model and a Multi-Objective Genetic Algorithm. *Lecture Notes in Artificial Intelligence* 3533, Springer, Berlin, pp 628-637
- Castells M (1996, 2000) *The Information Age : Economy, Society and Culture*. Blackwell, Cambridge, MA
- Cecchini A (1996) Urban modelling by means of cellular automata: generalised urban automata with the help on-line (Augh) model. *Environment and Planning B* **23**: 721-732

- Cecchini A (1999) New Information and Communication Technologies (NICT) and the Professions of Town and Regional planners. *Computers, Environment and Urban Systems* **23**
- Clarke KC, Gaydos LJ (1998) Loose-coupling a cellular automaton model and GIS: long-term urban growth predictions for San Francisco and Baltimore. *International Journal of Geographic Information Science* **12**: 699–714
- Coello CA, Van Veldhuizen DA, Lamont GB (2002) *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer Academic Publishers
- Colorni A, Ponti M, Brambilla M, Erba S (2003) Studio di fattibilità per una MetroTranVia nell'area di Como. Consorzio Poliedra, Milan Polytechnic
- Dalkey N (1972) An elementary cross impact model. *Technological Forecasting and Social Change* **3**: 3: 341–51
- Deb K, Jain S (2002) Running performance metrics for Evolutionary Multiobjective Optimization. KanGAL Report No. 2002004, Indian Institute of Technology, Kanpur
- Di Gregorio S, Serra R (1999) An empirical method for modelling and simulating some complex macroscopic phenomena by cellular automata. *Future Generation Computer Systems* **16**: 259–271
- Duperrin JC, Godet M (1973) Methode de hierarchisation des elements d'un systeme. Rapport economique du CEA
- Fonseca CM, Fleming PJ (1995) An overview of evolutionary algorithms in multiobjective optimisation. *Evolutionary Computation* **3**: 1: 1–16
- Font A (ed.) (2004) "L'explosió ee la ciutat / The Explosion of the City. COAC and Forum Universal de les Cultures Barcelona 2004
- Gigerenzer G, Todd PM (1999) *Simple heuristics that make us smart*. Oxford University Press
- Giron FJ, Rios S (1980) Quasi-Bayesian behaviour: A more realistic approach to decision making? In: Bernardo JM, DeGroot JH, Lindley DV, Smith AFM (eds) *Bayesian Statistics*. University Press, Valencia
- Godet M (1974) SMIC: a new cross impact method. *Futures* **8**: 4: 336–349
- Godet M (1984) *Prospective e planification stratégique*. CPE, Paris
- Godet M (1986) Introduction to 'la prospective': Seven key ideas and one scenario method. *Futures* **2**: 2: 134–57
- Quagliarella D (ed) (1998) *Genetic Algorithms and Evolution Strategy. Engineering and Computer Science: Recent Advances and Industrial Applications*. Wiley & Sons, Chichester
- Goldberg DE (1989) *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, Reading, MA
- Gordon TJ, Hayward H (1968) Initial experiments with the cross-impact method of forecasting. *Futures* **1**: 2: 100–116
- Hall P (1988) *Cities of tomorrow: an intellectual history of urban planning and design in the twentieth century*. Basil Blackwell, Oxford-New York
- Hanne T (2000) Global Multiobjective Optimization Using Evolutionary Algorithms. *Journal of Heuristics* **6**: 347–360
- Helmer O (1972) Cross-impact gaming. *Futures* **4**: 149–167
- Helmer O (1977) Problems in Future Research: Delphi and Causal Cross Impact Analysis. *Futures* **9**: 1: 17–31
- Holland JH (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI

- Honton EJ, Stacey GS, Millet SM (1985) *Future Scenarios: The BASICS Computational Method*. Battelle Columbus Division, Ohio
- Indovina F (ed.) (2005) *L'esplosione della Città*. Editrice Compositori, Bologna
- Kane J (1972) A primer for a new cross-impact language – KSIM. *Technological Forecasting and Social Change* **4**: 2: 149-167
- Krauth J, Duin H, Schimmel A (1998) A comparison of tools for strategic simulation and scenario generation with special emphasis on 'soft factors'. *Simulation Practice and Theory* **6**: 23-33
- Levi I (1989) *The Enterprise of Knowledge*. MIT Press, Cambridge, MA
- Linstone HA, Turoff M (eds.) (2002) *The Delphi Method: Techniques and Applications*, available at: <http://www.is.njit.edu/pubs/delphibook/index.html>
- Sadler B, Verheem R (1996) *Strategic Environmental Assessment: Status, Challenges and Future Directions*. Ministry on Housing, Spatial Planning and the Environment, Netherlands
- Sawaragi Y, Nakayama H, Tanino T (1985) *Theory of multiobjective optimisation*. Academic Press, Orlando, Florida
- Simon H (1981) *The science of the Artificial*. MIT Press, Cambridge, MA (1st edition: 1969)
- Srinivas N, Deb K (1995) Multiobjective function optimization using nondominated sorting genetic algorithms. *Evolutionary Computation* **2**: 3: 221-248
- Stover J, Gordon TJ (1978) Cross-impact analysis. In: Fowles J (ed) *Handbook of Futures Research*. Greenwood Press, Westport, CT, pp 119-128
- Takeyama M, Couclelis H (1997) Map dynamics: integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographic Information Science* **11**: 73-91
- Tobler W (1979) Cellular geography. In: Gale S, Olsson G (eds) *Philosophy in Geography*. Reidel, Dordrecht, pp 379-386
- de Toro F, Ortega J, Paechter B (2003) Parallel single front genetic algorithm: Performance analysis in a cluster system. *Proc. International Parallel and Distributed Processing Symposium*, Nice, France
- Turoff M (1972) An alternative approach to cross-impact analysis. *Technological Forecasting and Social Change* **3**: 3: 309-339
- Wagner D (1997) Cellular automata and geographic information systems. *Environment and Planning B* **24**: 219-234
- Walley P (1991) *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London
- Wellmann MP (1990) Fundamentals concepts of qualitative probabilistic networks. *Artificial Intelligence* **44**: 257-303
- Worsch T (1999) Simulations of cellular automata. *Future Generation Computer Systems* **16**: 157-170
- Wu F (1999) GIS-based simulation as an exploratory analysis for space-time processes. *Journal Geographic Systems* **1**: 199-218

Cities as Evolutionary Systems in Random Media¹

Leonid Bogachev

Department of Statistics, University of Leeds, UK

Abstract

The purpose of the paper is to discuss some potential applications of random media theory to urban modelling, with the emphasis on the intermittency phenomenon. The moment test of intermittency is explained using the model of continuous-time branching random walk on the integer lattice \mathbb{Z}^d with random branching rates. Statistical moments of the population density are studied using a Cauchy problem for the Anderson operator with random potential. The Feynman–Kac representation of the solution is discussed, and Lyapunov exponents responsible for the super-exponential growth of the moments are evaluated. The higher-order Lyapunov exponents are also obtained. The results suggest that the higher-order intermittency is reduced, in a sense, to that of the mean population density.

1 Introduction

There has been a long tradition in urban and regional science to approach the mathematical modelling of urban processes in the spirit of classical mathematical physics, by developing various deterministic (sometimes, quite sophisticated) models, aiming to capture the multi-component, dynamically variable nature of cities by augmenting the equations with suitable feedback mechanisms, akin to reaction-diffusion models or predator-prey models (see, e.g., Wilson 1981; Bracken and Tuckwell 1992, and further references therein). Although often useful in practical terms, the efficiency of such models proved to be limited to short spatio-temporal scales, hence there was the growing understanding among both theorists and practitioners that the deterministic approach was conceptually flawed, as it was missing some intrinsic, significant features of cities as agglomerates with formidably complex structure and intricate dynamics. Gradually, the simplistic paradigm of deterministic modelling has been superseded by the *complexity* paradigm, and in the last three decades or so it has become conventional to think of modern cities and regions as *complex spatial systems* (see Wilson 2000; Portugali 2000). More recently, ideas from the theory of *self-organization* proved to be useful in attempts to explain a familiar phenomenological picture of the city as an apparently “self-organizing” system, proceeding from agent-based models (see Portugali 2000).

¹ Research supported in part by DFG Grant 436RUS 113/722.

One significant feature characteristic to all such systems is the inherent *uncertainty* of their spatio-temporal dynamics, driven by various transport processes arising through interaction between “agents” at different structural levels and scales. In the urban context, this is exemplified by the competition between agents (individuals or organizations) via exchange of resources (e.g., land, labour, finances, goods, and information), which involves spatial mobility (transport networks and traffic, commuting flows, etc.) and population dynamics (demography, migration, etc.).

It is important to recognize that cities as systems are always non-closed due to interaction with the external medium (“environment”), which in turn may be random or fluctuating in its own right. This suggests that any conceptual urban model should involve appropriate random processes evolving in a *random medium*. One example of such processes is provided by *random walks in random environments*, where the usual random walk as a model of spatial motion of agents is modified so that the probabilities of jumps (or, in a continuous-time setting, the jumping rates) depend in a random fashion on the agent’s current location (Bogachev 2006). In situations where, in addition, there is a possibility of local reproduction or annihilation of agents (which may be modelled by *branching random walk*), it may be necessary to allow the branching probabilities (rates) to be space-dependent *and* random.

Modern theory of random media provides an adequate framework to model such systems (see, e.g., Zeldovich et al. 1988; Molchanov 1991). Typically, evolutionary processes in random media are driven by the two competing mechanisms: (1) self-averaging (via diffusion, dissipation, etc.) and (2) self-organization (via interaction, branching, local reactions, etc.). In the short term, when fluctuations of the medium are moderate enough, self-averaging (homogenization) usually prevails, allowing one to replace the complicated random medium by a much simpler *effective* non-random medium (see Kozlov 1985; Molchanov 1991). Loosely speaking, such a situation is analogous to the asymptotic normality in probability theory, where the combination of lots of small random factors generates a Gaussian distribution of fluctuations around non-random “mean”. In fact, the idea of homogenization underlies the classical mathematical physics and dates back to the pioneering work of Boltzmann (thermodynamics), Maxwell (electromagnetic dynamics) and Rayleigh (heat propagation); see a historical review and an account of modern development in Molchanov (1991).

However, over the past three decades or so, it has been observed that in many cases, due to statistically rare but extremely large fluctuations of the random environment, the interplay between self-averaging and self-organization may lead to anomalous effects that cannot be predicted or explained via the averaged approach, such as localization, intermittency, non-classical diffusion, etc. (see Zeldovich et al. 1988; Bouchaud and Georges 1990; Molchanov 1991, 1994, and further vast bibliography therein).

In spatio-temporal modelling of cities as complex systems, of particular interest is the *intermittency* phenomenon. This term, originally coined in the theory of hydrodynamic turbulence (see Monin and Yaglom 1975) and later on made popular in synergetics and nonlinear science (see Haken 1978; Mikhailov and Loskutov

1991), refers to the development of highly irregular structures in space and time, featured by a hierarchy of extremely high and sparse peaks (“spikes”) on a relatively low-profile background. Intermittency is typical for complex processes in many areas including chemical physics (strong centers in catalytic reactions), biology (formation of living organisms), ecology (propagation of species in the biosphere), genetics (gene expression on DNA microarrays), hydrodynamics (hierarchy of vortices in Kolmogorov–Obukhov’s theory of hydrodynamic turbulence), astrophysics (formation of galaxies, stars and planets), and economics (accumulation and redistribution of wealth and resources; modern globalization), to mention but a few (see Zeldovich et al. 1988; Bouchaud and Georges 1990; Molchanov 1991, and further references therein).

In the urban context, an example of intermittency is readily provided by cities as they stand, representing high, relatively localized peaks of the population density (cf. the hierarchy “villages – towns – cities – megacities – megalopolises”). Let us point out that first attempts to allow for stochastic factors in urban and regional models were made by collaborators within the *Brussels school* (see Allen et al. 1978; Allen and Sanglier 1979) who applied Prigogine’s ideas (see Nicolis and Prigogine 1977, 1989) to urban systems. A critical survey and discussion of the Brussels school’s approach can be found in Wilson (1981). In these works it was observed, mainly via numerical integration and computer simulations, that stochastic dynamics generates hierarchically-structured spatial patterns, but the role of randomness has largely remained obscure. Among more recent approaches, let us mention the thought-provoking work by Zanette and Manrubia (1997, 1998) and Manrubia and Zanette (1998) on application of intermittency ideas in urban modelling in an attempt to explain the Zipf law of the Pareto-type distribution of the size of large cities. An interesting example of intermittency has been recently reported by Giardina and Bouchaud (2003) for an agent-based market model.

The main purpose of the present paper is to draw a wider attention of urban and regional scientists to possible anomalies caused by the indispensable randomness of the environment. In particular, we aim to explain the powerful yet simple moment test of intermittency, proposed by Zeldovich et al. (1988) and developed further by Gärtner and Molchanov (1990) and Molchanov (1991, 1994). Heuristically, the qualitative picture of intermittency suggests that, due to high peaks, intermittent spatial structures are characterized by an anomalous ratio between consecutive statistical moments of the evolving field (so that, e.g., the second-order moment grows much faster than the square of the first-order one, etc.). We shall illustrate this picture using a simple model of branching random walk with random birth-and-death rates.

The paper is organized as follows. In Section 2 we describe our basic model of continuous-time branching random walk in d dimensions. In Section 3, we consider the mean population size m_1 in a fixed (*quenched*) branching medium and derive Kolmogorov’s equation for m_1 as a function of time and the initial position of the ancestor. The Feynman–Kac representation of the solution, which plays the crucial role in our analysis, is explained in Section 4. In Section 5 the long-term behaviour of m_1 is discussed, using the so-called *annealed* moments $\langle m_1^p \rangle$ obtained by further averaging over the medium. In particular, the Lyapunov exponents

are computed explicitly in the Weibull case. It follows that the moments $\langle m_1^p \rangle$ grow progressively, and in Section 6 we explain why this is an evidence of strong intermittency of m_1 . In Section 7, these results are extended to higher-order moments of the population size, which prove to be intermittent as well. In Section 8 we note a remarkable identity between the Lyapunov exponents of various orders, which suggests that the higher-order intermittency is in a sense reduced to the first-order intermittency. Finally, concluding remarks are made in Section 9.

2 The Model

In a simplified setting, we may think of an evolutionary spatial system (e.g., a city) as being driven by certain “agents” which move randomly in space and may interact locally with the environment (but not with each other). To be more specific, we suppose that the spatial motion is of type of diffusion, which will be modelled via continuous-time simple symmetric random walk on the integer lattice \mathbb{Z}^d ($d \geq 1$) with *jumping rate* $\kappa > 0$. The latter means that during a small amount of time $h \rightarrow 0$, the walk currently at site $x \in \mathbb{Z}^d$ can jump to an adjacent site x' ($|x' - x| = 1$) with probability $(2d)^{-1}\kappa h + o(h)$ (see Fig. 1 left). The remaining option, of course, is to stay at the current location over time h , which has probability $1 - \kappa h + o(h)$. (Here and below, the notation $y = o(h)$ means that $y = y(h)$ is much smaller than h , that is, $y(h)/h \rightarrow 0$ as $h \rightarrow 0$.) As is well known from the standard theory of Markov chains (see, e.g., Gihman and Skorohod 1975), an equivalent description of the random walk’s local dynamics is that it spends at each site an exponentially distributed random time (with mean $1/\kappa$), thereafter jumping to one of the $2d$ adjacent sites chosen at random, with equal probabilities $(2d)^{-1}$.

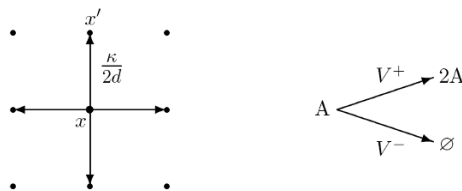


Fig. 1. Left: Random walk transition ($d = 2$). Right: Branching transitions

Furthermore, we suppose that the local variability of the population is governed by a birth-and-death process, whereby an agent may be either replaced by two new agents (at *reproduction rate* V^+) or removed from the system (at *annihilation rate* V^-). Possible branching transitions of an agent at each locus are schematically shown in Fig. 1 (right). It is assumed that the branching rates are variable in space: $V^\pm = V^\pm(x)$ ($x \in \mathbb{Z}^d$), and random (hence *random branching medium*). The proba-

bility distribution of the branching medium will be denoted by \mathbb{P} , and expectation with respect to \mathbb{P} by the angular brackets, $\langle \cdot \rangle$.

Thus, given a realization ω of the random medium (i.e., a collection of the branching rates $V^\pm(\cdot)$), an agent, currently at site $x \in \mathbb{Z}^d$, during a small time $h \rightarrow 0$ is replaced by two descendants with probability $V^+(x)h + o(h)$, dies with probability $V^-(x)h + o(h)$, or else stays intact, with probability $1 - V^+(x)h - V^-(x)h + o(h)$. Similarly to the random walk's description above, this implies that the random time until branching transformation has exponential distribution with parameter $V^+(x) + V^-(x)$, and when it expires the agent either splits into two or dies, with probabilities proportional to $V^+(x)$ or $V^-(x)$, respectively.

As usual, it is assumed that the mechanisms of random walk and branching act independently of each other, which amounts to saying that exponential random "clocks" measuring time until jumping and branching are statistically independent. Therefore, the dynamics of the population ensemble proceeds on the "microscopic" time scale according to the following rule. If at time t an agent is located at site x , then during a small subsequent amount of time $h > 0$, with probability $(2d)^{-1}\kappa h + o(h)$ it can jump to an adjacent site x' ($|x' - x| = 1$), with probability $V^+(x)h + o(h)$ it is replaced by two agents at the point x , or otherwise, with probability $V^-(x)h + o(h)$ the agent dies. Accordingly, with probability $1 - \kappa h - (V^+(x) + V^-(x))h + o(h)$ the agent experiences no transformations during the whole time h . The following table contains the list of all "elementary" transitions (up to probabilities of order of $h \rightarrow 0$):

Table 1.

Transition	Probability
Jump to adjacent site	$(2d)^{-1}\kappa h + o(h)$
Split into two	$V^+(x)h + o(h)$
Die	$V^-(x)h + o(h)$
Stay intact	$1 - [(2d)^{-1}\kappa + V^+(x) + V^-(x)] h + o(h)$

Newly born agents evolve by the same rule, independently of the other agents and the past history. We also impose the initial condition by assuming that at time $t = 0$ there is a single agent in the system located at some point $x \in \mathbb{Z}^d$.

In what follows, we denote by \mathbb{P}_x^ω the *quenched* probability law of this process, and by \mathbb{E}_x^ω the corresponding expectation. (The label ω refers to a fixed realization of the branching medium $V^\pm(\cdot)$, whereas the superscript x indicates the initial position of the single original agent.)

3 First-Order Quenched Moment

Let us denote by $X_t(y)$ the number of agents at time $t \geq 0$ at site $y \in \mathbb{Z}^d$, and consider also total population size $X_t := \sum_y X_t(y)$. Since at time $t = 0$ there is one agent in the system placed at point $x \in \mathbb{Z}^d$, we note that

$$X_0 = 1, \quad X_0(y) = \delta_x(y), \quad (1)$$

where $\delta_x(y)$ is the Kronecker symbol taking values 1 and 0 according as $x = y$ or $x \neq y$, respectively.

Fixing a realization ω of the branching medium, let us consider the *quenched* statistical moments of the first order:

$$m_1(t, x, y) := \mathbf{E}_x^\omega X_t(y), \quad m_1(t, x) := \mathbf{E}_x^\omega X_t.$$

Backward Kolmogorov's equations for these functions can be derived in the usual way, by decomposing with respect to possible transitions during the initial time $h \rightarrow 0$.

For example, for the mean population size $m_1(t, x)$ we can write, using Table 1:

$$\begin{aligned} m_1(t+h, x) = & \frac{\kappa h}{2d} \sum_{|x'-x|=1} m_1(t, x') + 2m_1(t, x)V^+(x)h + \\ & + \left[1 - \kappa h - V^+(x)h - V^-(x)h\right] m_1(t, x) + o(h) \end{aligned} \quad (2)$$

Indeed, the sum over the nearest neighbours x' in Eq. 2 accounts for possible jumps from x to x' , which amounts to re-starting the process at point x' , while the second term in Eq. 2 represents the possibility of branching, so that at time h we have two agents at x and hence the total mean population size after subsequent time t will be double the $m_1(t, x)$. Subtracting $m_1(t, x)$ from both parts of Eq. 2, dividing by h and passing to the limit as $h \rightarrow 0$, we arrive at the equation

$$\begin{cases} \partial_t m_1 = \kappa \Delta m_1 + V(x)m_1, \\ m_1(0, x) \equiv 1 \end{cases} \quad (3)$$

(for the initial condition, see Eq. 1). Here $\partial_t = \partial/\partial t$ is the time derivative, Δ is the discrete Laplacian acting on functions $\psi : \mathbb{Z}^d \rightarrow \mathbb{R}$ as

$$\Delta \psi(x) = \frac{1}{2d} \sum_{|x'-x|=1} \psi(x') - \psi(x)$$

and the function $V(x)$, called the *random potential*, is given by

$$V(x) := V^+(x) - V^-(x), \quad x \in \mathbb{Z}^d \quad (4)$$

Remark. By the symmetry of the operator $H = \kappa \Delta + V(x)$ (which implies time reversibility of the system), the solution $m_1(t, x)$ of the problem (3) may be interpreted as the mean number of agents at site x at time t , given that initially there is a uniform distribution of agents, $X(x) \equiv 1$ (i.e., one agent per site).

Similarly, for the local first-order moment $m_1 = m_1(t, x, y)$ we obtain the same differential equation as in Eq. 3 but with a different (localized) initial condition:

$$\begin{cases} \partial_t m_1 = \kappa \Delta m_1 + V(x)m_1 \\ m_1(0, x, y) = \delta_y(x) \end{cases} \quad (5)$$

Thus, the first-order quenched moments $m_1(t, x)$ and $m_1(t, x, y)$ are solutions to the Cauchy problems (3) and (5), respectively, which involves the celebrated Anderson operator $H = \kappa \Delta + V(x)$ (see Molchanov 1991, 1994). As explained in Gärtner and Molchanov (1990), the intermittency properties of the solution are deeply related to the localization properties of the spectrum of the operator H . In our analysis of intermittency, we will not use this analogy.

Qualitatively speaking, the problem (3) describes the mean flow of agents in the presence of random sources (points where $V(x) > 0$) and random sinks (points where $V(x) < 0$). Two competing effects are present (cf. Section 1): the diffusion mechanism represented by the Laplace operator Δ , and the branching mechanism represented by the random potential V . The diffusion tends to make the population density m_1 flat, while the potential makes it irregular by either inhibiting ($V < 0$) or enhancing ($V > 0$) the population growth.

4 The Feynman–Kac Formula

In this section, we discuss the Feynman–Kac formula, which furnishes a useful probabilistic representation of the solutions to the Cauchy problems (3) and (5). For the sake of technical simplicity, from now on we assume that the pairs $V^\pm(x)$ are independent and identically distributed (i.i.d.) for all $x \in \mathbb{Z}^d$, and in particular the potential $V(x) = V^+(x) - V^-(x)$ is also i.i.d. Then, under certain moment conditions on the potential V (which are automatically satisfied for i.i.d. potentials with Weibull tails to be considered below), the solutions $m_1(t, x)$, $m_1(t, x, y)$ admit the following Feynman–Kac representation (see Gärtner and Molchanov 1990; Molchanov 1991):

$$\begin{aligned} m_1(t, x) &= E_x \exp \left(\int_0^t V(x_s) ds \right), \\ m_1(t, x, y) &= E_x \exp \left(\int_0^t V(x_s) ds \right) \delta_y(x_t) \end{aligned} \quad (6)$$

Here E_x denotes expectation with respect to an auxiliary simple symmetric random walk x_t started at x (with the same jumping rate κ). Formulas (6) are in line with the “Lagrangian approach”, where instead of looking at agents visiting a given point, one follows the motion of a tagged agent. The statistical weight $\exp(\int_0^t V(x_s) ds)$ characterizes the “variation of mass” resulting from the interaction with the medium (Molchanov 1991).

Let us sketch the proof of formula (6) for $m_1(t, x)$. Supposing that the function m_1 is given by Eq. 6, let us verify that it satisfies the differential Eq. 3 (obviously, Eq. 6 implies the initial condition in Eq. 3). Using the Markov property (“lack of memory”) and the local structure of transitions of the random walk x_t , we have

$$\begin{aligned}
 m_1(t+h, x) &= \mathbb{E}_x \exp \left(\int_0^{t+h} V(x_s) ds \right) \\
 &= \mathbb{E}_x \exp \left(\int_0^h V(x_s) ds \right) \left[\mathbb{E}_{x_h} \exp \left(\int_0^{t+h} V(x_s) ds \right) \right] \\
 &= \mathbb{E}_x \exp \left(\int_0^{t+h} V(x_s) ds \right) m_1(t, x_h) \\
 &= \mathbb{E}_x \left(1 + \int_0^h V(x_s) ds + o(h) \right) m_1(t, x_h) \\
 &= (1 - \kappa h)(1 + V(x)h + o(h))m_1(t, x_h) + (1 + O(h))(m_1(t, x_h) - m_1(t, x)) \\
 &= m_1(t, x) + (V(x) - \kappa)hm_1(t, x) + \frac{\kappa h}{2d} \sum_{|x' - x| = 1} m_1(t, x') + o(h)
 \end{aligned}$$

Taking $m_1(t, x)$ to the left-hand side and dividing by $h \rightarrow 0$, we get Eq. 3.

5 Lyapunov Exponents for m_1

From the Feynman–Kac representation (Eq. 6), it is clear that the long-term behaviour of the solution is determined by large values of the potential $V(\cdot)$. Hence, the structure of the upper tail of the probability distribution of the potential is of principal importance.

In this work, we shall consider the case of Weibull-type tails:

$$\mathbb{P} \{V(x) > v\} \sim \exp(-cv^\alpha) \quad (v \rightarrow \infty), \quad (7)$$

with some $\alpha > 1$, $c > 0$. For instance, normal distribution $N(0, \sigma^2)$ belongs to this class, with $\alpha = 2$.

Recall that the moment functions $m_1(t, x)$, $m_1(t, x, y)$ depend on the random medium through the random potential $V(\cdot)$, and are therefore random variables. To investigate their asymptotic behaviour, let us consider the annealed statistical moments

$$\langle m_1(t, x)^p \rangle, \langle m_1(t, x, y)^p \rangle \quad (p = 1, 2, \dots)$$

where $\langle \cdot \rangle$ denotes expectation with respect to the environmental distribution \mathbb{P} .

The following result is essentially due to Gärtner & Molchanov (1990) (see also Zeldovich et al. 1988 for an earlier version and Albeverio et al. (2000) for the part with $m_1(t, x, y)$).

Theorem 1. Assume that the random potential $V(\cdot)$ is i.i.d. and has a Weibull tail (7) with $\alpha > 1$. Set $\alpha' := \frac{\alpha}{\alpha - 1} > 1$. Then the following limits (moment Lyapunov exponents) exist for both $m_1(t, x)$ and $m_1(t, x, y)$:

$$\lambda_p := \lim_{t \rightarrow \infty} \frac{\log \langle m_1^p \rangle}{t^{\alpha'}} = \gamma(\alpha, c) p^{\alpha'} \quad (p = 1, 2, \dots), \tag{8}$$

where

$$\gamma(\alpha, c) = (\alpha - 1) \alpha^{-\alpha'} c^{-1/(\alpha-1)}. \tag{9}$$

Note that the right-hand side of Eq. 8 does not depend on the starting point x , which is obviously due to the medium’s homogeneity. More striking is the fact that the Lyapunov exponents for the local moments $m_1(t, x, y)$ coincide with those for the total moments $m_1(t, x)$ (and thus are independent of the point y as well). This shows that in the long run, all sites $y \in \mathbb{Z}^d$, no matter how far from the starting point x , make the same contribution to the growth of the annealed mean population size.

Let us sketch the proof of Theorem 1 for the total moment $m_1(t, x)$. Consider the cumulant generating function

$$G(t) := \log \langle e^{tV(x)} \rangle$$

(since $V(\cdot)$ is i.i.d., the right-hand side does not depend on x). For a Weibull distribution (7) one can show that

$$G(t) \sim \gamma(\alpha, c) t^{\alpha'}, \quad t \rightarrow \infty, \tag{10}$$

where $\gamma(\alpha, c)$ is defined in Eq. 9. For example, if the distribution is given precisely by

$$\mathbb{P} \{V(x) > v\} = \exp(-cv^\alpha) \quad (v \geq 0),$$

then integration by parts gives

$$\langle e^{tV(x)} \rangle = \int_0^\infty e^{tv} d(1 - e^{-cv^\alpha}) = - \int_0^\infty e^{tv} d(e^{-cv^\alpha}) = t \int_0^\infty e^{tv - cv^\alpha} dv \tag{11}$$

By the Laplace asymptotic method, logarithmic asymptotics of the integral in Eq. 11 is determined by the maximum of the function $g(v) = tv - c v^\alpha$. Solving the equation

$$g'(v) = t - c \alpha v^{\alpha-1} = 0$$

yields that the maximum is attained at $v_0 = (t/c\alpha)^{1/(\alpha-1)}$. As a result,

$$\begin{aligned} \log \langle e^{tV(x)} \rangle &\sim g(v_0) = \\ &= t^{\alpha/(\alpha-1)} c^{-1/(\alpha-1)} \left(\alpha^{-1/(\alpha-1)} - \alpha^{-\alpha/(\alpha-1)} \right) = \\ &= \gamma(\alpha, c) t^{\alpha'} \end{aligned}$$

in accord with Eq. 10.

We are now able to derive a suitable *lower bound*. Let us keep in Eq. 6 only *one* path of the random walk x_s , namely the one that stays at its initial point x up until time t . Since the sojourn time at x exceeds t with probability $e^{-\kappa t}$, we get

$$m_1(t, x) \geq e^{-\kappa t} \cdot e^{V(x)t}$$

hence

$$\langle m_1(t, x)^p \rangle \geq e^{-p\kappa t} \langle e^{-ptV(x)} \rangle = e^{-p\kappa t + G(pt)} \tag{12}$$

To obtain an upper bound asymptotically equivalent to the lower one, we apply Jensen’s inequality and use Fubini’s theorem as follows:

$$\begin{aligned} \langle m_1(t, x)^p \rangle &= \left\langle \left(\mathbf{E}_x \left[\exp \left\{ \int_0^t V(x_s) ds \right\} \right] \right)^p \right\rangle \leq \\ &\leq \left\langle \left(\mathbf{E}_x \left[\exp \left\{ p \int_0^t V(x_s) ds \right\} \right] \right) \right\rangle \leq \\ &\leq \mathbf{E}_x \left\langle \frac{1}{t} \int_0^t e^{ptV(x_s)} ds \right\rangle = \\ &= \frac{1}{t} \int_0^t \mathbf{E}_x \langle e^{ptV(x_s)} \rangle ds = \\ &= \langle e^{ptV(x_s)} \rangle = e^{G(pt)} \end{aligned} \tag{13}$$

The combination of the estimates of Eqs. 12 and 13 yields

$$\frac{-\kappa t + G(pt)}{t^{\alpha'}} \leq \frac{\log \langle m_1(t, x)^p \rangle}{t^{\alpha'}} \leq \frac{G(pt)}{t^{\alpha'}}$$

and the result of Theorem 1 follows in view of Eq. 10.

In conclusion of this section, let us point out that Gärtner and Molchanov (1990) have also evaluated the *quenched* Lyapunov exponent

$$\tilde{\lambda} := \lim_{t \rightarrow \infty} \frac{\log m_1}{t(\log t)^{1/\alpha}} = \left(\frac{d}{c} \right)^{1/\alpha} \tag{14}$$

which exists \mathbf{IP} -almost surely (i.e., with probability 1). Let us explain this result briefly, following Zeldovich et al. (1988). The crucial observation is that the main contribution to the expectation in Eq. 6 for large values of t is provided by an *optimal* path of the random walk x_t , which travels at a distance of order of t from the origin x to reach the maximum value of the potential

$$V_0 = \max \{ V(y) : |y - x| \leq t \}$$

(rather than by a *typical* path with $|x_t - x| \approx t^{1/2}$). To estimate a typical value of the maximum V_0 , note that the probability $p = \mathbb{P}\{V(y) > v\}$ to exceed a high level $v \rightarrow \infty$ at a given point y equals $p = \mathbb{P}\{V(y) > v\} \approx \exp(-cv^\alpha) \rightarrow 0$. Neglecting the unlikely possibility that the maximum value V_0 is attained at more than one point and using that there are about $N \approx t^d$ integer points in the ball of radius t , from the relation $\mathbb{P}\{V_0 > v\} \approx Np \approx 1$ we get the estimate

$$V_0 \approx \left(\frac{d \log t}{c}\right)^{1/\alpha} \quad (t \rightarrow \infty).$$

On the other hand, the probability of the optimal path can be shown to be asymptotically of order of $\exp(-c_1 t)$. Inserting these estimates into formula (6), we obtain

$$\log m_1 \approx -c_1 t + t \left(\frac{d \log t}{c}\right)^{1/\alpha} \sim t \left(\frac{d \log t}{c}\right)^{1/\alpha} \quad (t \rightarrow \infty),$$

and Eq. 14 follows.

Heuristically, the results (8) and (14) imply a super-exponential temporal growth as $t \rightarrow 1$:

$$\langle m_1^p \rangle \approx \exp(\lambda_{1,p} t^{\alpha'}), \quad m_1 \approx \exp(\tilde{\lambda} t (\log t)^{1/\alpha}), \quad (15)$$

where “ \approx ” denotes the logarithmic equivalence. Note that the time scale for the annealed moments $\langle m_1^p \rangle$ grows much faster than the time scale for the moment m_1 itself. This can be explained by the presence of extremely high peaks in the spatial distribution of the population density, which is an indicator of intermittency. In the next section, we will address the intermittency issue in more detail.

6 Intermittency of m_1

The main qualitative inference from Theorem 1 is that we deal here with a strongly intermittent structure. To see why, we note from Eq. 3 that the *specific Lyapunov exponents* λ_p/p are given by

$$\frac{\lambda_p}{p} = \gamma(\alpha, c) p^{\alpha'-1}. \quad (16)$$

Note that the Lyapunov inequality, for any $p < q$ we have

$$\langle m_1^p \rangle^{1/p} \leq \langle m_1^q \rangle^{1/q},$$

so that

$$\frac{\log \langle m_1^p \rangle}{p} \leq \frac{\log \langle m_1^q \rangle}{q}.$$

This implies that the specific exponents λ_p/p are always *non-decreasing* in p ,

$$\frac{\lambda_p}{p} \leq \frac{\lambda_q}{q} \quad (p < q)$$

Moreover, from Eq. 16 it follows that λ_p/p is *strictly increasing* as a function of p , since $\alpha' > 1$:

$$\lambda_1 < \frac{\lambda_2}{2} < \frac{\lambda_3}{3} < \dots$$

which implies a progressive temporal growth of the moments:

$$\langle m_1^2 \rangle \gg \langle m_1 \rangle^2, \quad \langle m_1^3 \rangle \gg \langle m_1^2 \rangle^{3/2}, \text{ etc.}$$

As was first observed by Zeldovich et al. (1988) and argued more carefully by Gärtner and Molchanov (1990), this is an indication of strong intermittency in the spatial distribution of m_1 , and it is our aim to explain their argument here.

For definiteness, we shall speak of the total moment $m_1(t, x)$. Let us pick a number r_1 such that $\lambda_1 < r_1 < \lambda_2/2$, and consider the (random) level sets on the lattice \mathbb{Z}^d on which the magnitude of m_1 exceeds an exponentially high level:

$$\Gamma_{r_1}(t) := \left\{ x \in \mathbb{Z}^d : m_1(t, x) > \exp(r_1 t^{\alpha'}) \right\}.$$

The spatial density of points in the set $\Gamma_{r_1}(t)$ can be determined as the limit

$$\rho_{r_1}(t) := \lim_{R \rightarrow \infty} \frac{\#(\Gamma_{r_1}(t) \cap B_R)}{\#(B_R)}$$

where $B_R = \{x \in \mathbb{Z}^d : |x| < R\}$ is the “ball” of radius $R \rightarrow \infty$ and $\#(B)$ is the number of points $x \in \mathbb{Z}^d$ contained in the set $B \subset \mathbb{Z}^d$. By general results from ergodic theory (Birkhoff–Khinchin’s theorem), the density $\rho_{r_1}(t)$ coincides with the probability that the set $\Gamma_{r_1}(t)$ contains a given site x :

$$\rho_{r_1}(t) = \mathbb{P} \left\{ x \in \Gamma_{r_1}(t) \right\} = \mathbb{P} \left\{ m_1(t, x) > \exp(r_1 t^{\alpha'}) \right\}$$

Applying Chebyshev’s inequality and using Eq. 15, we see that

$$\begin{aligned} \mathbb{P} \left\{ m_1(t, x) > \exp(r_1 t^{\alpha'}) \right\} &\leq \langle m_1(t, x) \rangle \cdot \exp(-r_1 t^{\alpha'}) \\ &\approx \exp(\lambda_1 t^{\alpha'}) \cdot \exp(-r_1 t^{\alpha'}) \\ &= \exp\left(- (r_1 - \lambda_1) t^{\alpha'}\right) \rightarrow 0 \quad (t \rightarrow \infty) \end{aligned}$$

hence the spatial density of $\Gamma_{r_1}(t)$ is exponentially small. On the other hand, it is easy to see that the contribution to the second-order moment $\langle m_1(t, x)^2 \rangle$ from points *outside* $\Gamma_{r_1}(t)$ is exponentially negligible:

$$\begin{aligned} \langle m_1(t, x)^2 I\{x \notin \Gamma_{r_1}(t)\} \rangle &= \langle m_1(t, x)^2 I\{m_1(t, x) \leq \exp(r_1 t^{\alpha'})\} \rangle \\ &\leq \exp(2r_1 t^{\alpha'}) \\ &\ll \exp(\lambda_2 t^{\alpha'}) \approx \langle m_1^2(t, x) \rangle \end{aligned}$$

(here $I\{A\}$ denotes the indicator of event A). That is to say, the second-order moment $\langle m_1^2 \rangle$ is basically formed by high “overshoots” of $m_1(t, x)$ that occur on the set $\Gamma_{r_1}(t)$

In a similar way, choosing r_2 such that

$$\frac{\lambda_2}{2} < r_2 < \frac{\lambda_3}{3},$$

one can show that the third-order moment m_3 is essentially formed on the level set

$$\Gamma_{r_2}(t) = \{x \in \mathbb{Z}^d : m_1(t, x) > \exp(\lambda_2 t^{\alpha'})\} \subset \Gamma_{r_1}(t),$$

and so on. Thus, there is a hierarchical sequence of spatially sparse level sets

$$\Gamma_{r_1}(t) \supset \Gamma_{r_2}(t) \supset \dots \supset \Gamma_{r_p}(t) \supset \dots \quad (r_1 < r_2 < \dots < r_p < \dots)$$

such that each moment $\langle m_1^p \rangle$ is generated by overshoots of the field $m_1(t, x)$ occurring on the respective set $\Gamma_{r_p}(t)$, which corresponds to the heuristic picture of intermittency.

7 Higher-Order Quenched Moments

In the previous sections we have seen that the moment approach proves quite efficient in the intermittency analysis of the first-order quenched moments $m_1(t, x) = \mathbf{E}_x^\omega X_t$, $m_1(t, x, y) = \mathbf{E}_x^\omega X_t(y)$ using their annealed moments $\langle m_1^p \rangle$.

The same method can be applied to studying the higher-order quenched moments

$$m_n(t, x) := \mathbf{E}_x^\omega X_t^n, \quad m_n(t, x, y) := \mathbf{E}_x^\omega X_t^n(y)$$

(see Albeverio et al. 2000). From the qualitative picture of intermittency explained above, it is natural to expect that the functions m_n will appear strongly intermittent as well, the situation probably getting even more irregular as compared to the case of m_1 . Our aim in this section is to present the results concerning m_n .

For the sake of clarity, let us assume that the annihilation rate V equals zero, so that the branching medium is completely determined by the random potential $V(\cdot) = V^+(\cdot) \geq 0$. One can show (see Albeverio et al. 2000) that the functions m_n satisfy the chain of evolution equations

$$\begin{cases} \partial_t m_n = \kappa \Delta m_n + V(x)m_n + V(x)f_n \\ m_n(0, x) \equiv 1, \quad m_n(0, x, y) = \delta_y(x) \end{cases} \quad n = 1, 2, \dots \tag{17}$$

where

$$f_n \equiv f_n [m_1, \dots, m_n] := \sum_{i=1}^{n-1} \binom{n}{i} m_i m_{n-i}$$

and $\binom{n}{i} = \frac{n!}{i!(n-i)!}$ are the binomial coefficients.

The initial-value problem (Eq. 17) can be viewed as an *inhomogeneous* Cauchy problem for the Anderson operator $H = \kappa \Delta + V(x)$ (cf. Eqs. 3 and 5). Similarly to the homogeneous case, the solution of this problem admits a representation of the Feynman–Kac type (see details in Albeverio et al. 2000); for example, $m_n(t, x)$ can be represented in the form (cf. Eq. 6)

$$m_n(t, x) = m_1(t, x) + \mathbb{E}_x \int_0^t \exp \left(\int_0^s V(x_u) du \right) V(x_s) f_n(t-s, x_s) ds$$

The higher-order *annealed* Lyapunov exponents, specifying the rate of super-exponential growth of $\langle m_n^p \rangle$, are given by the following theorem (Albeverio et al. 2000).

Theorem 2. *Under the same assumptions as in Theorem 1 above (that is, i.i.d. $V(\cdot)$, Weibull tail (Eq. 7) with $\alpha > 1$), for both $m_n(t, x)$ and $m_n(t, x, y)$ it holds*

$$\lambda_{n,p} := \lim_{t \rightarrow \infty} \frac{\log \langle m_n^p \rangle}{t^{\alpha'}} = \gamma(\alpha, c)(np)^{\alpha'} \quad (n = 1, 2, \dots, p = 1, 2, \dots) \tag{18}$$

where $\alpha' := \alpha/(\alpha-1) > 1$ and the constant $\gamma(\alpha, c)$ is the same as in Theorem 1 (see Eq. 9).

Note that for $n = 1$, we have $\lambda_{1,p} \equiv \lambda_p$ (see Theorem 1), and formulas (18) are reduced to the expressions (8).

The *quenched* Lyapunov exponents of m_n were evaluated by Molchanov (1996):

$$\tilde{\lambda}_n := \lim_{t \rightarrow \infty} \frac{\log m_n}{t(\log t)^{1/\alpha}} = n \left(\frac{d}{c} \right)^{1/\alpha} \quad (P - \text{a.s.}), \tag{19}$$

which is a generalization of formula (14).

Comparing the results (18) and (19), we note that the time scales for the annealed and quenched asymptotics of m_n are quite different, the former growing much faster than the latter. This situation is analogous to that for the first-order moment m_1 (see Section 5), and again is an indication of intermittency. This is confirmed by progressive growth (in both p and n) of the specific higher-order annealed Lyapunov exponents,

$$\frac{\lambda_{n,p}}{p} = \gamma(\alpha, c) n^{\alpha'} p^{\alpha'-1}$$

$$\frac{\lambda_{n,p}}{n} = \gamma(\alpha, c) n^{\alpha'-1} p^{\alpha'}$$

8 Adequacy of Moments

The principal qualitative conclusion from the previous analysis may seem somewhat disappointing, as it appears that the annealed moments are *nonrepresentative* (Molchanov 1991). Indeed, the moments of orders up to p leave out information about higher peaks, which are only “visible” in the moments of orders $p' > p$.

However, the mere comparison of the results (18) and (19) reveals the following remarkable relation between the various Lyapunov exponents responsible for the superexponential growth of the annealed moments:

$$\lambda_{n,p} = \lambda_{1,np} \quad (n = 1, 2, \dots, p = 1, 2, \dots) . \tag{20}$$

Heuristically, the identity (20) implies that

$$\langle m_n^p \rangle \approx \langle m_1^{np} \rangle \quad (t \rightarrow \infty)$$

Thus, relative to the first-order moment m_1 , the higher-order moments m_n behave (almost) in a “regular” fashion. This observation makes it plausible that intermittency of m_n (and possibly of the underlying population density X_t as well) essentially amounts to that of the first-order moment m_1 (cf. Molchanov 1996, where a similar observation was made).

This conjecture is further confirmed by the comparison of the quenched Lyapunov exponents (14) and (19):

$$\tilde{\lambda}_n = n \tilde{\lambda}_1 \quad (P - \text{a.s.})$$

so that, with probability 1,

$$m_n \approx m_1^n \quad (t \rightarrow \infty)$$

Note that the overall (annealed) probability measure \wp_x , defining the population process X_t in random branching medium V^\pm , can be symbolically represented by the total probability formula:

$$\wp_x(\cdot) = \int_{\{\omega\}} P_x^\omega(\cdot) \mathbb{P}(d\omega) = \langle P_x^\omega(\cdot) \rangle$$

We conjecture that the following is true:

Adequacy Principle. *The ratio*

$$\rho(t) := \frac{X_t}{m_1(t, x)}$$

is bounded in \wp -probability from above and below, uniformly in $t \geq 0$, i.e., for any $\varepsilon > 0$ there exists $\delta > 0$ such that for all $t \geq 0$

$$\wp \left\{ \delta \leq \rho(t) \leq \delta^{-1} \right\} \geq 1 - \varepsilon.$$

Note that from the Adequacy Principle it would follow that

$$\frac{\log X_t}{\log m_1(t, x)} \xrightarrow{\wp} 1 \quad (t \rightarrow \infty),$$

provided that

$$m_1(t, x) \xrightarrow{\mathbb{P}} \infty \quad (t \rightarrow \infty).$$

In turn, this would imply the equality of the (quenched) Lyapunov exponents for X_t and $m_1(t, x)$: if for some deterministic scale function $A(t) \uparrow \infty$ one has

$$\lim_{t \rightarrow \infty} \frac{\log m_1(t, x)}{A(t)} = 1 \quad (\mathbb{P}\text{-a.s.}),$$

then it follows that

$$\lim_{t \rightarrow \infty} \frac{\log X_t}{A(t)} = 1 \quad (\wp\text{-a.s.}).$$

We will address these issues elsewhere.

9 Concluding Remarks

In this paper, we have focused on the notion of intermittency, referring to the formation of irregular spatio-temporal structures in complex systems, with the view of potential applications in urban and regional modelling. We have demonstrated that the model of branching random walk with *random* branching rates may produce a strongly pronounced intermittency of the (quenched) mean population distribution m_1 in space and time. In so doing, progressive growth of annealed Lyapunov exponents $\lambda_{1,p}$, with respect to the moment order p , proves to be a simple yet efficient test of intermittency.

Let us emphasize that, due to randomness of the medium, intermittency arises here in a *linear* problem of the form $\partial_t u = Hu$, although the actual mechanism of intermittency is a nonlinear dependence of the solution, through the Feynman–Kac representation, on the random coefficients of the equation.

We have also considered the higher-order quenched moments m_n , for which the similar technique is available. In particular, the higher-order Lyapunov exponents $\lambda_{n,p}$ may be evaluated, confirming the intermittent nature of the process. A remarkable identity between the various Lyapunov exponents, $\lambda_{n,p} = \lambda_{1,np}$, suggests that the annealed moments $\langle m_n^p \rangle$ grow in a “regular” fashion relative to the first-order moment.

Therefore, although intermittent structures are very hard to deal with statistically, the main work to be done relates to the first-order moment only, which seems to be a good news (relative to the “bad” news of intermittency!).

Needless to say, the model we have discussed is far too idealized and simplified to be of any practical use. There are also further mathematical issues to be addressed in order to gain a better understanding of intermittency, including its fine spatial structure. In more realistic models, proper modifications should be made to allow for other types of probability distribution of the potential, and also for finiteness of space and time, spatial capacity constraints, temporal variability and spatial correlations of the medium, and interaction between the agents.

Some of these issues are still open, while some other have already been addressed in the literature, for example:

- continuous-space models (see Gärtner & König 2000);
- dependence of the spatial structure of peaks on the distribution tail of the potential (see Gärtner and Molchanov 1998; Biskup and König 2001);
- time-dependent models with short temporal correlations (see Carmona & Molchanov 1994);
- spatial correlation structure of the quenched moment (see Gärtner and den Hollander 1999).

However, our principal aim here has been to provide a quick introduction to intermittency as part of random media theory, in order to show that random media models may be of great relevance and value when it comes to understanding and explanation of various anomalous effects. The main qualitative message from our analysis to applied scientists and practitioners, and in particular to urban modelers, is that in an intermittent situation the usual statistical moments may not be representative of the actual process under consideration, so greater caution is needed when using the standard sampling techniques (based on averaging etc.). One practical advice based on the moment test is that should the higher-order moments, as functions of time, appear to be growing progressively in their order (which may be manifested, for instance, through “instability” or “divergence” of statistical and numerical estimates), then this may be a precursor of intermittency.

Acknowledgements

Partial support from the organizers of this Conference is gratefully acknowledged. I am pleased to express my deep gratitude to Stanislav Molchanov for numerous illuminating conversations over the years on many issues in random media theory and to Sergio Albeverio for stimulating collaboration. I have also appreciated interesting discussions on various aspects in urban applications with Peter Allen, Mike Batty, Itzhak Benenson, Mark Birkin, Volker Jentsch, Juval Portugali, and Damian Zanette. Special thanks are due to Sir Alan Wilson for his seminal books promoting good use of mathematics in urban and regional modelling.

References

- Albeverio S, Bogachev LV, Molchanov SA, Yarovaya EB (2000) Annealed moment Lyapunov exponents for a branching random walk in a homogeneous random branching environment. *Markov Process. Relat. Fields* **6**: 473–516
- Allen PM, Deneubourg JL, Sanglier M, Boon F, de Palma A (1978) The dynamics of urban evolution. 1: Interurban evolution; 2: Intraurban evolution. Final Report to the US Department of Transportation, Washington DC
- Allen PM, Sanglier M (1979) A dynamic model of growth in a central place system. *Geograph. Anal.* **11**: 256–272
- Biskup M, König W (2001) Long-time tails in the parabolic Anderson model with bounded potential. *Ann. Probab.* **29**: 636–682
- Bogachev LV (2006) Random walks in random environments. In: Françoise JP, Naber G, Tsou ST (eds) *Encyclopedia of Mathematical Physics*. Elsevier, Oxford, vol 4, pp 353–371
- Bouchaud J-P, Georges A (1990) Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Phys. Rep.* **195**: 4-5: 127–293
- Bracken AJ, Tuckwell HC (1992) Simple mathematical models for urban growth. *Proc. Roy. Soc. London Ser. A* **438**: 171–181
- Carmona RA, Molchanov SA (1994) Parabolic Anderson Problem and Intermittency. (Memoirs of the American Mathematical Society; No. 518.) American Math. Society, Providence RI
- Gärtner J, den Hollander F (1999) Correlation structure of intermittency in the parabolic Anderson problem. *Probab. Theory Relat. Fields* **114**: 1–54
- Gärtner J, König W (2000) Moment asymptotics for the continuous parabolic Anderson model. *Ann. Appl. Probab.* **10**: 192–217
- Gärtner J, Molchanov SA (1990) Parabolic problems for the Anderson model. I: Intermittency and related topics. *Comm. Math. Phys.* **132**, 613–655.
- Gärtner J, Molchanov SA (1998) Parabolic problems for the Anderson model. II: Second-order asymptotics and structure of high peaks. *Probab. Theory Relat. Fields* **111**, 17–55.
- Giardina I, Bouchaud J-P (2003) Bubbles, crashes and intermittency in agent based market models. *Eur. Phys. J. B, Condens. Matter Phys.* **31**: 421–437
- Gihman II; Skorohod AV (1975) *The Theory of Stochastic Processes 2*. Springer, Berlin
- Haken H (1978) *Synergetics: An Introduction*. Springer, Berlin.
- Kozlov SM (1985) The method of averaging and walks in inhomogeneous environments. *Russian Math. Surveys* **40**: 2: 73–145
- Manrubia SC, Zanette DH (1998) Intermittency model for urban development. *Phys. Rev. E* **58**: 295–302
- Mikhailov AS, Loskutov AY (1991) *Foundations of Synergetics II: Complex Patterns*. Springer, Heidelberg
- Molchanov SA (1991) Ideas in the theory of random media. *Acta Appl. Math.* **22**: 139–282
- Molchanov SA (1994) Lectures on random media. In: Bakry D, Gill RD, Molchanov SA (eds) *Lectures on Probability Theory. Ecole d'Eté de Probabilités de Saint-Flour XXII-1992*. Lecture Notes Math. **1581**: 242–411

- Molchanov SA (1996) Reaction-diffusion equations in the random media: Localization and intermittency. In: Funaki T, Woyczynski WA (eds) *Nonlinear Stochastic PDEs. Hydrodynamic Limit and Burgers' Turbulence*. Springer, Berlin, pp. 81–109
- Monin AS, Yaglom AM (1975) *Statistical Fluid Mechanics: Mechanics of Turbulence 2*. MIT, Cambridge MA
- Nicolis G, Prigogine I (1977) *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations*. Wiley, New York
- Nicolis G, Prigogine I (1989) *Exploring Complexity: An Introduction*. W. H. Freeman, New York
- Portugali J (2000) *Self-Organization and the City*. Springer, Berlin
- Wilson AG (1981) *Catastrophe Theory and Bifurcation: Applications to Urban and Regional Systems*. Groom Helm, London
- Wilson AG (2000) *Complex Spatial Systems: The Modern Foundations of Urban and Regional Analysis*. Pearson Education, Harlow
- Zanette DH, Manrubia SC (1997) Role of intermittency in urban development: A model of large-scale city formation. *Phys. Rev. Lett.* **79**: 523-526
- Zanette DH, Manrubia SC (1998) Reply to: Comment on "Role of intermittency in urban development: A model of large-scale city formation". *Phys. Rev. Lett.* **80**: 4831
- Zeldovich YaB, Molchanov SA, Ruzmaikin AA, Sokoloff DD (1988) Intermittency, diffusion and generation in a nonstationary random medium. *Sov. Sci. Rev., Sect. C: Math. Phys. Rev.* **7**: 1–110

Grilling the Grid: a Non-Ultimate (Nor Objective) Report on the Configurational Approach to Urban Phenomena

Valerio Cutini

Department of Civil Engineering, University of Pisa, Italy

Abstract

This paper is a report on the configurational theory, the unconventional approach to urban phenomena that was introduced by Bill Hillier in the mid '80s and still attracts and stimulates researchers all over the world. It won't be exhaustive, since it can't but neglect most of the several techniques that, on the common configurational basis, have been worked out so far. It won't be ultimate, since at present researchers are still working hard and carrying on developments on the matter. It won't be objective either, since the stated purpose of the paper is the introduction and the discussion of a new, original method, the MaPPA, and its placement as the logical terminus of decades of studies and experimentations.

But, all in all, the underlying purpose of the paper is to outline the usefulness and the reliability of such approach, to highlight benefits and limits of the several techniques, as well as to figure possible lines of improvement and development of the presented methods.

1 Introduction

This paper is around the urban grid, that is the weave of spaces and paths which compose each urban settlement. And such a theme is suggested by the special importance a significant current of urban modelling assigns it. The grid – how it has to be read, how to be analysed, how to be questioned and (literally) grilled - will hence be the sense of the paper. The treatment can't be expected extensive nor all-inclusive (that's not its purpose), but rather to be a wide roundup on the techniques which have been introduced in the last 20 years on the common conceptual basis: the real aim is showing the usefulness of the methods, but even more showing their respective faults, which actually limit their concrete use in town planning. A secondary aim is to justify, on the basis of such faults, the opportunity of introducing further, different techniques, whose features will be sketched below.

“By giving shape and form to our material world, architecture structures the system of space in which we live and move. In that it does so, it has a direct relation – rather than a merely symbolic one – to social life, since it provides the ma-

terial preconditions for the patterns of movement, encounter and avoidance which are the material realization – as well as sometimes the generator – of social relations” (Hillier and Hanson 1984). With these words, Bill Hillier and Julienne Hanson state in 1984 the foundations of the configurational theory. Here we find, clearly expressed, the actual sense of the change of view over urban phenomena that the configurational theory does introduce: it’s no longer the network of interactions between couples of located activities what determines the distribution of movement along the urban paths and the layout of land uses and land values; no longer what happens within the buildings (houses, shops, offices, schools, factories, other activities) will be assumed as the primary cause of what happens outside. What matters is the urban space or, better, the whole set of streets and squares which are actually available for movement and constitute the urban grid. The grid will then assumed as the primary element of urban phenomena, that’s to say the distribution of movement and the location of activities along the urban paths (or, what’s better saying, their vocation for housing them). As a consequence of this change of view, urban space assumes a fundamental role in urban analysis. In the traditional urban modelling the space remains in the background, as the passive element which merely provides the spatial impedance between couples of activities; in configurational analysis the space is on the centre of the stage, as it’s assumed as the active input which determines the distribution of movement and activities all over the grid. Several effects can be easily drawn out of this new approach. First, the capability of working at a smaller scale, when applied to analyse limited portion of the urban settlement (even up to the architectural level), while traditional quantitative models are typically best suited to wide scale analysis (Lee 1973). Moreover, a fresh attention over the morphological aspects (the form and the geometry of streets, squares, blocks, even buildings), which were widely neglected or even ignored in interaction modelling (Lee 1973). Thanks to these effects, configurational analysis appear suitable to bridge the gap which traditionally divides urban design (appointed at projecting the transformation of the material structure of a settlement, mainly at a local scale) from territorial analysis (aimed at understanding and simulating its effects on both material and immaterial variables, chiefly at a wider scale), and let those two fields work together and concur in the development of settlements (Hillier 1996a).

It’s worth noticing that the consideration of space as an essential element in the use of an urban settlement is actually anything but new; since the most remote ages, the geometric morphology of an urban site was constantly assumed as the most prominent variable in the distribution of activities and land uses: wall defences, market places, political, religious and administrative buildings, residences were located on the basis of such geographic and geometric evaluations. Nor is original the attempt to investigate the use of a settlement by means of a topological approach based on the paths along its streets: as a well known example, we can mention the logic game that Euler proposed (Fig. 1) to determine the way of moving in the town of Königsberg, touching all its neighbourhoods and passing only once through each of the existing seven bridges over the river Pregel (Euler 1736).

Yet, the ancient Eulerian game, as aimed at connecting the movement in an urban settlement with the position of the existing paths, could at the most be considered as a remote archetype of configurational approach: what's really fresh in the configurational theory is the possibility of using the spatial pattern of a settlement as an input to determine, for each single part of it, a set of numeric variables suitable to describe its specific features with respect to all the others, that's to say its configurational state; and, furthermore, the consideration of this configurational state as the primary element in the actual use of each part of the settlement.

Space hence does matter, or, even, in Hillier's words, *space is the machine*. But how space has to be assumed, measured, evaluated, that's all another question, which researchers in the last two decades have been working on.

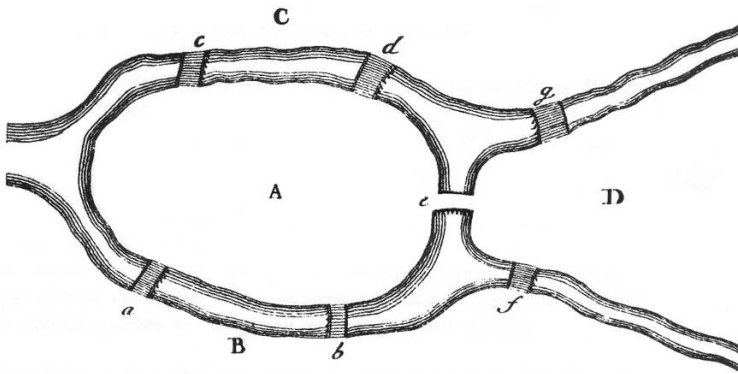


Fig. 1.

Fig. 1. The Eulerian problem of the bridges of Königsberg. A planimetric layout (Euler 1736)

2 The Configurational Approach: Conceptual Basis

As described above, the configurational approach assumes the urban grid as the primary element in the genesis of urban processes; and that assumption is based on the hypothesis of the existence of the natural movement, which is differently arranged along the paths of the settlement. It's then possible connect an urban grid to a specific distribution of natural movement, which appears favouring some locations (streets and squares highly crowded with movement) and disliking some others, characterised by lower density of movement. Urban activities will then likely locate in the most favoured places, so as to intercept a greater through movement. Such location, on the other hand, will then produce some further movement, attracted by the existing activities; on its turn, this movement will induce the location of other activities. And so on, in an exponential process which

enhances the positional advantages of the most appealing locations attracting there more and more activities and producing more and more movement. It's easy to recognise at the root of this exponential increase of movement and activities the primary element of the urban grid, while the activities appear as the mere multiplier of its effects.

Before going into the operational methods for appraising the grid and its configuration, it's worth specifying the relations between the configurational approach and the traditional urban theories and modelling. The change, it's worth saying, is radical but not really heretical: each single movement is still to be considered as the material result of an interaction between an origin activity and a destination one. Nonetheless, should we assume all the activities (and hence all the possible origins and destinations) as uniformly and capillarily distributed, the resulting interaction movement (that we call attracted movement in that it directly results from the attraction interaction between a couplers of activities) would be appraised as a through movement, global function of the grid and not depending on the presence and on the position of the singles activities. That movement, depending on the mere urban grid, is called natural movement and its presence is the basis hypothesis of the configurational approach. In the assumed case, natural movement and attracted movement will then coincide. Obviously, the capillary and uniform distribution of activities is anything but a realistic hypothesis, since activities are far from being homogeneous and uniformly located. If their actual location reproduces the distribution of configurational indices, they will multiply the effects of the grid on the arrangement of movement; in the (likely) case of a different distribution, natural movement and attracted one will diverge, as a consequence of the divergence between grid configuration and land use. In real cases, not necessarily natural movement ought to be preponderant with respect to the attracted one; it may even be smaller, but still remain the primary one, since it derives from the grid and from its inner configuration (Hillier et al. 1993).

Such aspect can be clearly understood if we conceptualise an urban grid as an element determining some potential movement along its paths: this potentiality can (or otherwise cannot) be realized as a consequence of the actual presence of buildings, activities and regulations. In other words, the grid works as "*a mechanism for generating contacts*" (Hillier 1996b) between the activities located in its areas, that's to say as a structure aimed at optimizing the movement and to maximise the consequent interactions and contacts. Each part of the grid is hence provided with a specific vocation towards the contacts between activities; such vocation can be followed by the actual land use (as it happens in most cases), or otherwise it can clash with it. As an example of this possible contrast, we can mention the imposed location of activities into poorly appealing areas, or the presence of regulations which assign the ineligibility of some urban areas for building development. Some discordance between configuration and land use (that's to say some clash between natural movement and attracted one), far from disproving the dependence of urban phenomena on the configuration of the grid, are instead at attesting the poor utilization of the potential movement economies which the grid makes available (and that the configurational analysis can easily point out) (Hillier 1996b).

After clearing the relevance of the change of view and the importance of the urban space in this different approach, it's necessary to specify how this space ought to be appraised, or, in better words, to precisely define the sense and the meaning of the term "configuration". In a configurational approach, each space in a settlement is appraised on the base of its connections to all the others, and in particular the relation between any two spaces depends on their relations to all the other spaces in the system. On such basis, to analyse the configuration of an urban grid means to define the characteristics of each of its spaces as they result from the relations connecting it to all the other parts of the grid; down to the facts, it means to define, for each single element of the system, a set of numeric values correspondent to its configurational indices; these indices, all together, describe the configurational state of the system. Are then the relations connecting each element to all the others, rather than its geometric or morphologic features what determine the configuration of the system and the configurational values of each of its elements. And just the measure of these relations, rather than other aspects, what can be properly said the configuration of the urban grid. On this regards, a little example will be useful to clear the sense of configuration and the differences with respect to the structure and the morphology of a settlement. In Fig. 2a is represented a planimetric layout, composed of 9 square cells, mutually connected by means of "doors"; in order to highlight the configurational state of the settlement it's useful to convert this layout into a graph, representing the cells with nodes and the connections (if existing) with segments (Fig. 3a). Let's now introduce a minimal change (Fig. 2b), opening some doors between couples of cells and closing some others.

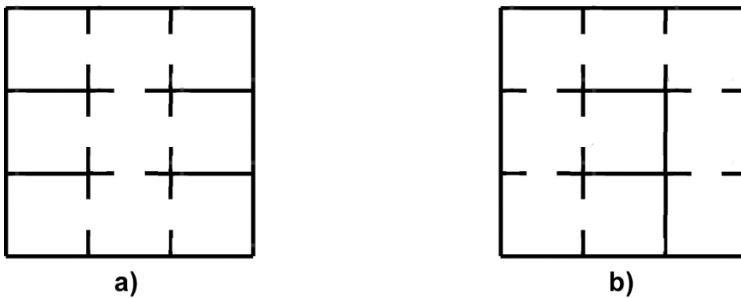


Fig. 2. A small example. Planimetric layout with 9 urban cells

In Fig. 3b is clearly shown how the relational state of the settlement appears radically transformed, while its structural and morphological aspects (Fig. 2b) remain unaltered.

If, then, as we've seen so far, the relations between the spaces of the urban grid, rather than their respective morphological and structural features, are the essential basis for the configurational approach, it's worth specifying that the fundamental relation the Hillierian theory takes into account is the visual connection between spaces.

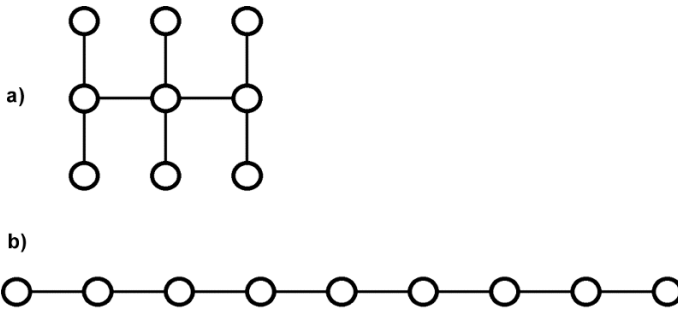


Fig. 3. Graph representation of the relational state of the layouts of Fig. 2

In fact, the way of appraising the spatial impedance is topologic, and its unit of measurement is essentially the visual prospective connecting each couples of urban places. Such an assumption reveals a debt towards the Gestaltpsychology: the way an urban settlement is actually used depends on the way do people move along its paths, and the way people move, on its turn, depends on the way they perceive and understand the space. It's then the visual perception what primarily indicates how to move within the urban space, and hence how the urban space can be conveniently used. The relevance of the geometric and morphological consistency of a settlement weakens with respect to the imageability (Lynch 1960) of its paths, and down to the facts that feature depends on the possibility of connecting its parts by means a limited number of viewsheds.

In determining the spatial impedance between urban locations, the configurational approach does hence substitute the metric distance with a measure of its perceptive appraisal; and such appraisal, roughly speaking, consists of the number of the interposed viewsheds. Such an approach is based on the assumption that each movement within the urban grid follows the visual appraisal of the final destination, by means of the sequence of the intermediate destinations scattered along the shortest path. And the number of viewsheds between two places, rather than the geometric consistency of the distance or the necessary time or money to cover it, is then assumed as decisive in the choice of the path. What essentially matters, roughly speaking, is not the material distance but the frequency of the necessary changes of visual perspective from the origin to the destination. The mental map of a settlement appears hence as reducing the spatial impedance between places in visual connection, and, on the contrary, widening their distance if it results from a sequence of viewsheds: what one can see is nearer than what is actually near but not visible.

The deterrence the tortuous paths determine in urban movement is well exemplified in a famous *music hall* sketch: in Dublin (where the historic centre is typically characterised by a tortuous and tangled grid) a man asks for an indication to reach Old Kent Road, and gets an answer like this: "Well, if I were you I wouldn't start from here" (Hamer 1999).

3 The Configurational Approach: Techniques and Methods

On the conceptual bases so far described, two aspects are still to be cleared and specified. First, the way the urban grid has to be discretised, that is transformed into a set of discrete elements, mutually interconnected by means of a specific interrelation. Such operation is necessary to analyse the urban space according to a systemic approach, so as to provide each single element with state variables.

Then, it's worth specifying which variables are to be assumed as the most suitable quantitative parameters, in order to describe the configurational features of those elements: thanks to such parameters, we'll be able to construct a hierarchy of the elements of the grid with respect to their capability of attracting movement.

Around these two aspects, the configurational theory appears anything but a unitary and monolithic vision: different approaches have in fact so far arisen, so that it's possible to ascribe them to two methods of analysing the urban space, the so-called *linear analysis* and *visibility graph analysis*. Such methods share the fundamental bases of the configurational vision (as they have been briefly described so far), which can be summarized as follows:

- the main (when not exclusive) focus on the relations between the spatial elements of the grid, rather than on their functional, morphological or structural features;
- the assumption of the urban grid as the primary element of urban phenomena;
- the hypothesis of the existence of natural movement, depending only on the configuration of the grid;
- the role of visual perception in the spatial relations between urban spaces.

On the other hand, the several configurational techniques differ in the specific way of resolving the urban grid into separate and discontinuous elements. Before going into the specific description of the techniques, it's worth noticing that those differences, far from corresponding to a simple operational distinction, actually reflect a different vision of urban phenomena. In other words, it's not a question of mere techniques, but rather a matter of a different philosophy in the approach to urban phenomena.

4 Operational Techniques: the *Linear Analysis*

The so-called *linear analysis* is based on the assumption of the linear path as the fundamental element of the urban space. A pedestrian perceives the urban space by means of segments which correspond to his own visual perspectives and, on their bases, moves from any origin to any destination along a sequence of intermediate lines. Is then just the line the essential key for the comprehension of the phenomena depending on the distribution of movement in the urban grid (Hillier and Hanson 1984). down to the facts, the linear analysis decompose the urban grid

in a system, which consists of a set of intersected segments. The first step in such a construction is the convex map, which is defined as the set of the widest convex spaces covering all the urban grid, taken in their minimum number. Based on the convex map, it's then possible to construct the axial map, which is defined as the complex of all the longest segments that connect all the convex spaces of the convex map, taken in their minimum number. The notion of convex space doesn't need any particular specification: as it's well known, it's defined as a polygon composed by couples of point all included within itself. The space corresponding to the polygon of Fig. 4a is hence convex, while the one represented in Fig. 4b is not.

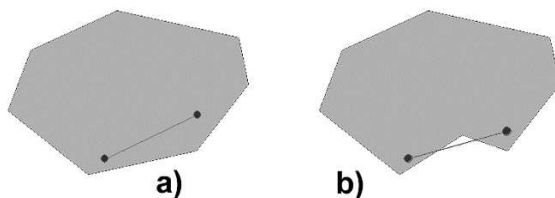


Fig. 4. Example of convex and non-convex spaces

Such a definition appears significant if we correlate the notion of connection segment between two points with the meaning of their reciprocal visual relation. From such view, a convex space can easily be interpreted as a portion of the urban grid characterised by perceptive unity. In other words, a convex space is composed of points in condition of mutual visibility., where each point can be seen from any other point within itself. Connecting the convex spaces by axial lines, the resulting axial map can then be seen as the network of the visual connections linking the single perceptive unities of the settlement. The pedestrian moves in the grid following the axial lines, invisible threads tightened between couples of convex spaces. In linear analysis, the meaning of visual perception (that takes shape in the viewshed) and the notion of movement (which follow it in the use of space) appear so correlate as to be actually undistinguishable. In order to better understand this kind of approach, it's here shown on a real urban case the process of construction of the system, here we present the small example of the historic centre of Pietrasanta (Fig. 5), so as to show how it's possible to derive from its urban grid (presented in figure 6a in a black on white representation, to focus on the real object of configurational analysis) the resulting convex map (Fig. 7) and then, on its base, the axial map (Fig. 8). That settlement was chosen for its small dimension, but also for some features which make it well fit for a clear highlighting of the various configurational techniques.

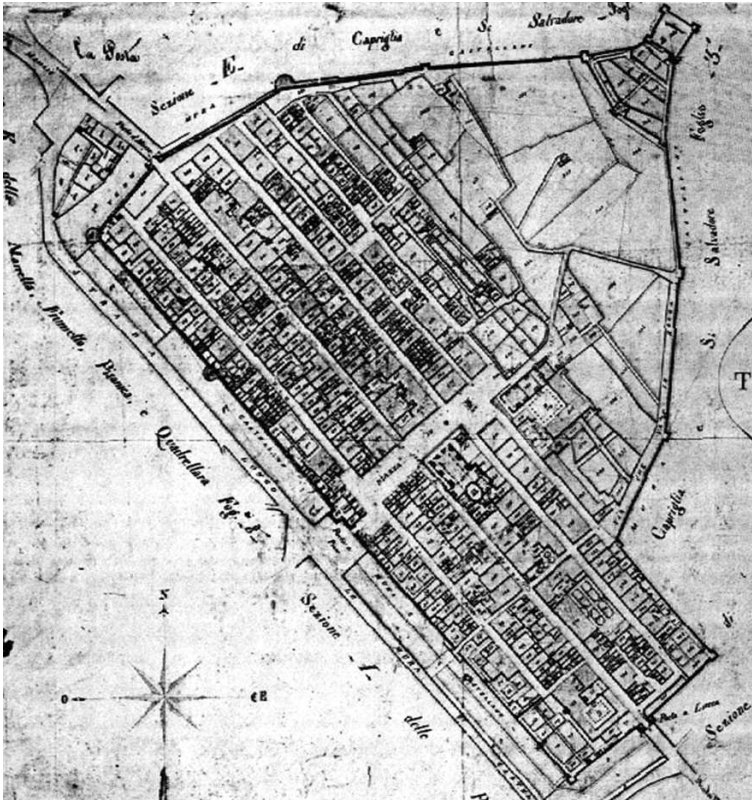


Fig. 5. The historic centre of Pietrasanta, as it appears represented in the Lorraine Cadastre (1824)

By means of the axial map, the urban grid is therefore converted into a set of discrete elements; in order to construct a system, it's then necessary to establish the relations which connect those elements. First, we introduce the belonging relation, by which it's possible to define its limits; such a role is assumed by the intersection between lines: only the intersected lines belong to the system. Roughly speaking, to fix such a belonging relation means to decide that only the convex spaces which result visually perceivable from other convex spaces are included within the system. Beside the belonging relation, we also need to define the structure relation, by which all the elements of the system are in mutual interaction.

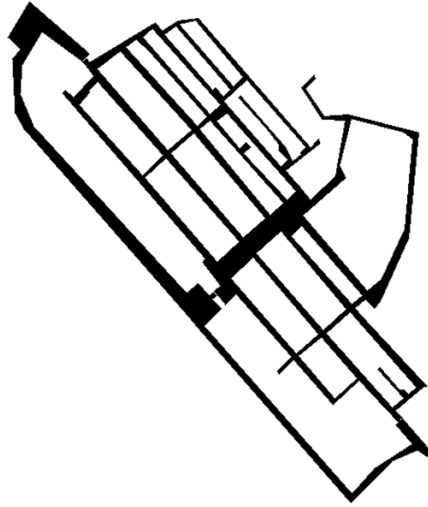


Fig. 6. The urban grid of the historic centre of Pietrasanta, in a black on white representation

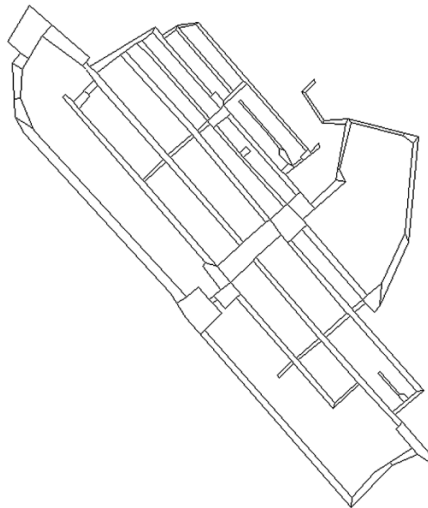


Fig. 7. The convex map of the historic centre of Pietrasanta

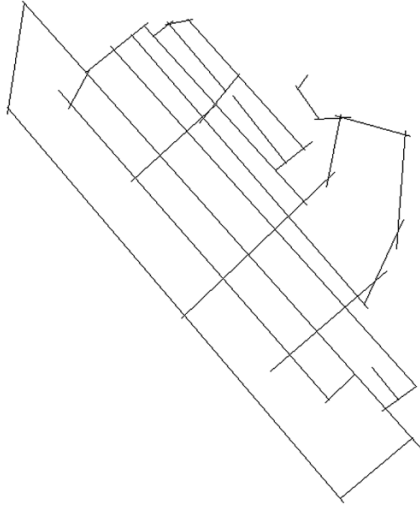


Fig. 8. The axial map of the historic centre of Pietrasanta

On such basis, it's then possible to provide each line of the system with a set of numeric variables, the configurational indices, that we previously did recognise as the state variables of the system. Here we present, briefly summarized, the most used and significant configurational indices (Hillier and Hanson 1984).

The connectivity of a line is defined as the number of lines that it's intersected with, or, what's the same thing, the number of lines which result unitarily deep with respect to it. In an axial map composed of k lines, the value of connectivity of a single line will then vary between 1 (minimum value, corresponding to a line connected to only another one) and $(k-1)$ (maximum value, corresponding to a line connected to all the other lines of the system). Down to the facts, a high value of connectivity characterizes a line which is provided with many views over other convex spaces.

Quite a different meaning is expressed by the control value: it represents the degree of control of a line with respect to the paths which involve it. Numerically, its value results from the sum of the inverse of the connectivity values of the connected lines, so as to vary from $(1/k-1)$ to $(k-1)$. The minimum threshold corresponds to a line with only a connection, whose connected line is on its turn connected to all the others; the maximum threshold corresponds to a line connected to all the other lines, when each other line has on its turn only a connection. Down to the facts, the control value reproduces the capability of a line to be seen from the connected ones as the possible terminal of any movement from and to themselves.

The integration value is by far the most significant (and most used) configurational index, and is defined as the mean depth of a line with respect to all the other lines of the axial map. That parameter allows to identify the most integrated line in the system and the most segregated one. In other words, the integration index is suitable to describe the mean accessibility of a line with respect to the whole sys-

tem, and since its value results from the consideration of the mere grid (without taking into account the presence, the position, the specific type and the consistency of the located activities), integration can be interpreted as a “pure accessibility”, in order to distinguish it from the traditional notion of accessibility. As it’s obvious, the numeric value of the integration index, according to such a definition, depends on the number of the lines of the system, that is on the actual dimension of the settlement, so that bigger grids will be meanly provided with higher integration values. Several normalization expressions have been introduced in order to cut off such dependence and to release the value of integration from the size of the urban grid, so to allow a direct comparison between settlements of different dimensions, or even to compare different consistencies of the same settlement due to different development projects.

The configurational analysis of the urban grid can be said worked out as each single line of the axial map results provided with its own set on numeric values for the desired variables. With reference to the most significant index, here we present in Fig. 9 an example of the distribution of its values in the already mentioned grid of Pietrasanta, the different values of integration are here represented by means of the thickness of the line: the thicker appears the segment, the more integrated is the correspondent segment (that is, of course, the lower is its numeric value).

It’s then possible to define a further configurational parameter, the local integration index, whose value is computed taking into account only the lines which lie within a topological circle around the considered line; generally, the assumed radius of the circle is 3, and the correspondent index is also called radius 3 integration. In Fig. 10, the distribution of R3 integration is represented on the same grid of Pietrasanta.

Due to the small dimension of the analysed settlement, obviously only few slight differences characterize the two presented distributions.

A further kind of linear analysis, which was introduced only few years ago (Turner 2001a), is the angular analysis. Unlike the more traditional linear analysis (which, to differentiate, is also called axial analysis), the angular one takes into account, beside the connection between couples of lines, also the respective interaction angle. The basis for this branch is the conviction (Sadalla and Montello 1989; Montello, 1991) that the spatial impedance of a change of visual perception gets bigger as the correspondent angle gets smaller: in other words, wide angles intersections ($> 90^\circ$) are less deterrent than small angles ones ($< 90^\circ$).

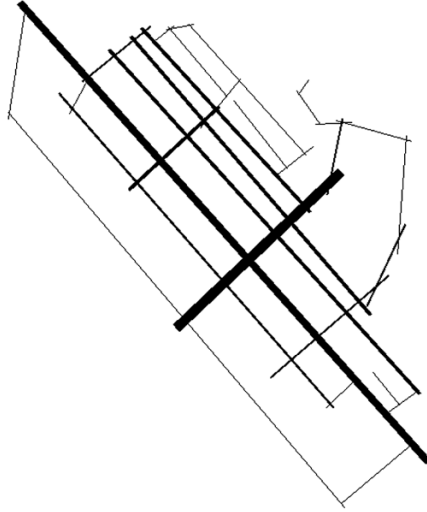


Fig. 9. The distribution of global integration in the urban grid of Pietrasanta. The value of the lines is proportional to the thickness of the correspondent segments

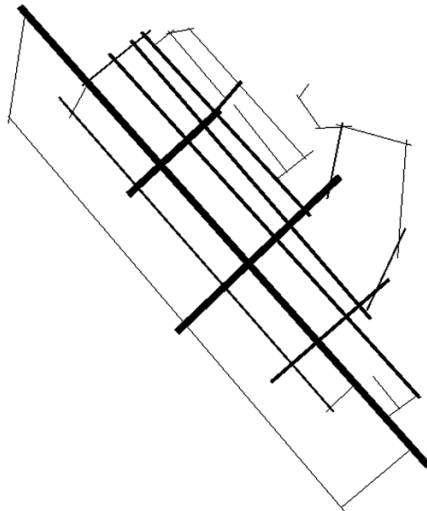


Fig. 10. The distribution of local (radius 3) integration in the urban grid of Pietrasanta. The value of the lines is proportional to the thickness of the correspondent segments

Wide angle intersections will then provide the depth between lines with a bigger contribute than the small angle ones, so that its value results from a weighted sum of the several angular steps. Analytically, such an assumption can be expressed as follows (Turner 2001a):

$$Md_i^\alpha = \sum_{j \in V(L)} d_{ij}^\alpha / \sum_{k \in E(L)} w_k$$

where d_{ij}^α is the shortest angular distance (that's to say the angular depth) between two lines, $V(L)$ is the set of all the lines of the axial map, $E(L)$ is the set of all the connection between lines and w_k the weight (function of the intersection angle) of those connections. As it can be easily understood, differences in the results between axial analysis and angular analysis will affect urban grids characterized by a wide range of angles in the intersection between lines. The urban case of Pietrasanta, which was presented above, as mainly composed of orthogonal intersections, is clearly poorly influenced, in the results of the analysis, by this methodological variant.

5 Operational Techniques: the *Visibility Graph Analysis*

In the late nineties another configurational approach to the analysis of urban settlements was introduced and named *visibility graph analysis*. This technique shares the same conceptual basis of linear analysis: the genesis of an urban settlement is primarily determined by the way its urban space is used, with particular reference to the distribution of movement along its paths. And, on its turn, this distribution of movement is determined by the way he visually perceives the space of the urban grid. And all that provides each space of the grid with a specific level of attractiveness towards movement and activities. What instead distinguishes the two approach is the way of decomposing the urban grid so as to obtain a system of interacting spatial elements. VGA is characterized by the assumption of the single point of the urban space as the minimum element of the grid (Turner et al. 2001): each single point is in fact assumed as the possible location of a pedestrian, who will move within the urban grid guided by the way he actually perceives other points, possible (final or intermediate) destinations of his movement.

From the convex space (composed by all the points in mutual visual connection) the focus passes to the isovist (composed by all the points visible from a single location) (Batty 2001); it's worth specifying that, unlike the convex spaces, the isovists are obviously intersected and partially overlapping, since several points are actually visible from different locations. Such an assumption simplify the operation of construction of the system, as it will be easily obtained covering the whole grid of the urban settlement with a mesh of points (called vertices) in mutual visual connection so as to obtain the so-called visibility graph: the visual connection between the couples of elements is then the belonging relation of the system, while the depth is once more assumed as its interaction relation. The only discretionary operation is such construction is the choice of the density of vertices: a denser mesh will provide more detailed results, while a less dense mesh allows a less complicated analysis but rougher results. Also the vertices of the VGA, just like the lines of linear analysis, are provided with a set of parameters, which are to be assumed as the variables of the configurational state of the system (Turner 2001b).

The neighbourhood size is the number of connected vertices; as strictly related to the isovist of a vertex, it corresponds to the extension of the urban area directly visible from a location. The neighbourhood size is the VGA correspondent of the connectivity in linear analysis and can be expressed as follows (Turner 2001b):

$$N_i = \{v_j : e_{ij} \in E\}$$

where e_{ij} is the connection between v_i and the neighbouring vertex v_j .

The clustering coefficient of a vertex is the ratio between the number of visual connection within its own isovist and the number of those which could in theory exist. This parameter identifies the level of convexity of the isovist of a vertex: it's clear that, if an isovist is convex, the clustering coefficient of its vertex would be 1. down to the facts, the clustering coefficient expresses the level of intervisibility of a vertex, that it how many vertices share its own isovist, or, what the same thing, how its isovist get transformed passing from a vertex to another one within itself. Analytically, the clustering coefficient can be expressed as follows (Turner 2001b):

$$\gamma_i = |E(\Gamma_i)| / k_i(k_i - 1)$$

where $E(\Gamma_i)$ is the set of the actually existing connections within the isovist and k_i its *neighbourhood size*.

Also the control value have a correspondent index in VGA and a similar definition, so that it can be analytically expressed as follows (Turner 2001b):

$$c_i = k_i / |U V(\Gamma_i) : v_j \in (\Gamma_i)|$$

The mean depth is defined as the mean number of (visual) steps one has to go through to reach all the other vertices of the visibility graph from a considered one. Once again, the mean depth hence can be regarded as an indicator of the (inverse) accessibility of the vertex. It's possible to use normalized expressions, so as to release its numeric value from the size of the grid, that is from the number of vertices of the visibility graph (what is highly aleatory since it's fixed by the operator).

As we've made above for the linear analysis, here we present in Fig. 11 the results of the visibility graph analysis of the urban grid of Pietrasanta, with reference to the distribution of global integration. Some comments will easily arise from the comparison of the results of the two different methods:

- from a general point of view, the distribution of the integration values in the two layouts appears fundamentally analogous, with the integration core located on the main longitudinal street and (especially) on the transversal one, which can be recognized as the strongest integrators.

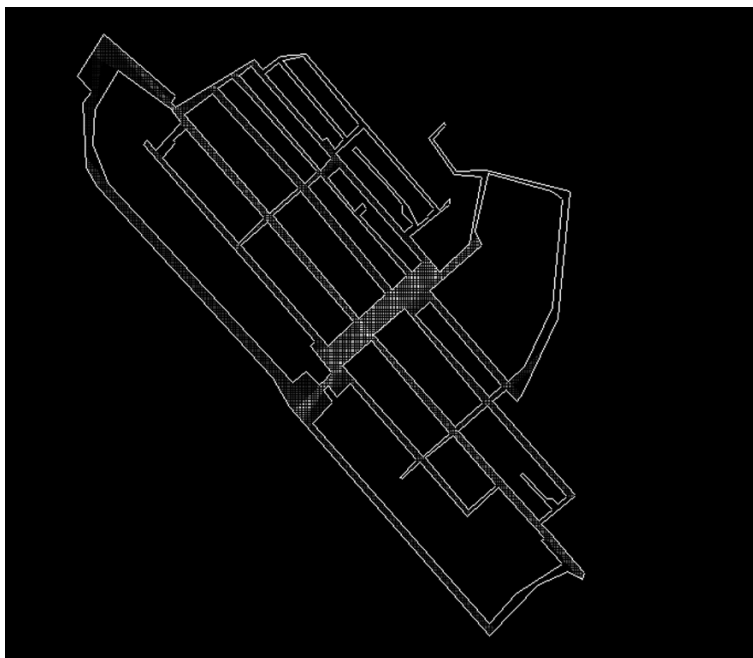


Fig. 11. The distribution of global integration in the urban grid of Pietrasanta as it results from VGA. In this representation, the integration value increases as the tone appears whiter

- looking more in details, we can clearly appreciate the variation of integration in correspondence of each single straight street, with different values along its length and specifically high values in the crossings;
- the VGA results provide detailed information about the distribution of integration within the wide open spaces, which simply do not appear in linear analysis;
- the numbers of elements in the two systems are dramatically different: 30 lines in the axial map versus more than 12.000 vertices in the visibility graph; that makes the VGA far more costly in computing power and processing time.

6 Operational Techniques: Potentials, Limits and Integrations

A brief comparison of the sketched techniques was already shown above, with reference to the case study of Pietrasanta. Few notes are useful to highlight the potential of such techniques, as well as to admit the respective limits.

Several studies attest a strong correlation of the configurational indices with the distribution of pedestrian movement (Cutini 1999a, 1999b) and with the actual density of located activities (Bortoli and Cutini 2001; Cutini 2001a). In Fig. 12 we present such correlation as it results from the analysis of the case studies of Siena

and Volterra. On such basis, configurational indices are recognized as suitable to reproduce the distribution of attractiveness towards both movement and activities, as the conceptual assumptions do actually assert. The configurational techniques can hence be used both to analyse urban settlements (in order to understand their inner geography and their genesis), and to support town planning in the location of activities (in order to make their presence to match with the actual spatial attractiveness). Moreover, the same techniques are suitable to guide the planner in arranging the location of monopolistic activities, whose location can even be chosen setting aside any economic evaluation, in order to enhance the centrality level of segregated areas (Cutini 2000).

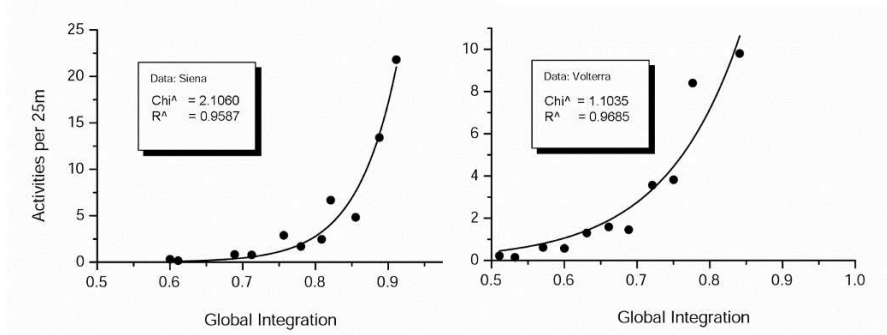


Fig. 12. Correlation analysis of integration value versus density of activities in the lines of the axial maps of Siena and Volterra

On the other side, the configurational techniques so far described appear affected by several faults which cannot be ignored. Talking of linear analysis, we can't but mention the following main limits:

- some subjectivity in the construction of the system, which can be affected by discretionary choices of the operator;
- the constancy of each index along the whole extension of a single line;
- the inability in taking into account the wide open spaces of the grid, which is reduced into a unidimensional system.

On its turn, also VGA shows some faults:

- the dimension of the system, which imposes complicated computing processing;
- the number of the elements of the system, whose variables are extremely difficult to export for the analysis and the use in other models or applications;
- the lacking of any correspondence between the elements of the system (merely scattered dots) and the spatial elements of the grid (streets, squares, blocks).

In order to overcome those limits, we've been proposing the introduction of some integrations that, while respecting the essential configurational basis, are aimed at favouring and extending their actual use in town planning.

VGA was enriched (Cutini 2003) with the introduction of a further configurational parameter, so-called interaction index, aimed at taking into account the capability of wide open space to be successfully used as meeting, gathering and interacting spaces, that's to say as squares. Such parameter results from the integration of several pre-existent configurational indices, and precisely:

$$K = 100 S C (1 - I)^\alpha$$

where S is the mean normalised neighbourhood size, C the mean clustering coefficient and I the normalized value of mean depth.

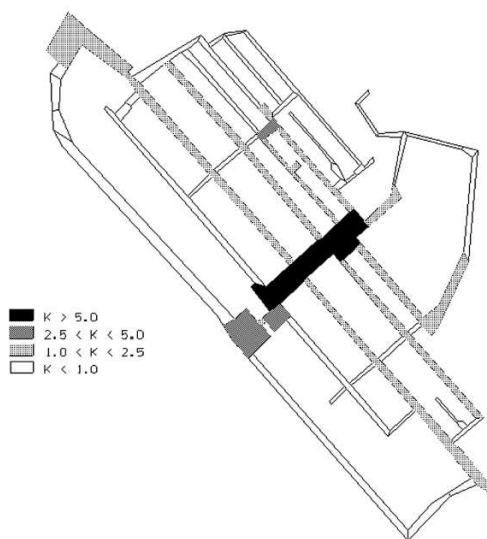


Fig. 13. Distribution of the interaction value in the convex spaces of Pietrasanta

Focusing on our case study, the distribution of the interaction value in the convex spaces of Pietrasanta is here represented in Fig. 13, which highlights with different tones of grey the most efficient squares on the settlement.

More recently, a different configurational technique was introduced (Cutini et al. 2004), proposing a new way of constructing the system. Named Ma.P.P.A. (Mark Points Parameter Analysis), the technique provides that the system shall be composed of the following elements:

- streets intersection points,
- direction change points,
- squares central points,
- access points to wide opens spaces,
- maximum distance points in a linear path,
- slope change points.

The assumption of those elements (the mark points) allows to overcome several faults of the pre-existent configurational techniques, that we've mentioned above; moreover, it's worth highlighting that the mark points can be automatically (and then objectively) extracted from a GIS, and that even the results of the processing of the system can easily be exported into a GIS, in order to better arrange and represent them, and especially use them in other applications (Cutini et al. 2004).

In Fig. 14 the distribution of integration in the mark points of the grid of Pietrasanta is represented by circles of different radius; as it can be easily seen by direct comparison with Figs. 9 and 11, such distribution appears somehow analogous to the ones seen so far, with the advantages we've sketched above.

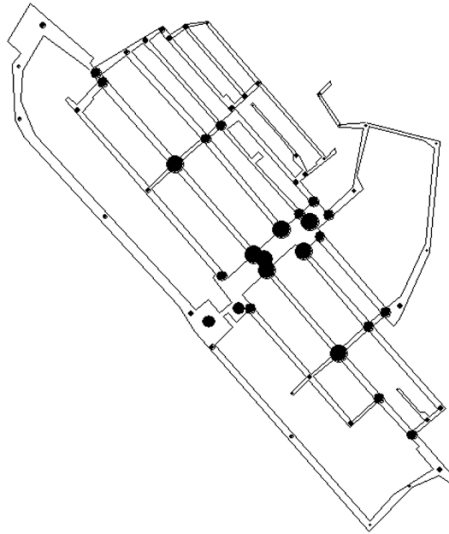


Fig. 14. The distribution of global integration in the mark points of the grid of Pietrasanta, processed by MAPP. The value is proportional to the radius of the correspondent circle

6 Conclusions

Space does matter, as we've seen. And space matters according to the relations which connect each single space to each other. Such connections do transform the urban grid as we know it (streets, squares, etc.) into a network of spatial relations between elements. This paper was about which elements are to be assumed as the elements of the system, that is how the grid has to be grilled in order to extract information on the use of the space. From this point of view, the roundup on the presented techniques can be seen as a search for a more and more automated and objective transformation of the grid, and for a more and more detailed resulting information. On such basis, this effort seems to come to an end: Ma.P.P.A. auto-

mates both the importing of the inputs and the exporting of the results, and those results are undoubtedly well detailed. Yet, such result is far from being ultimate. Other themes are to be faced, other question to be solved in order to allow a wider use of configurational techniques in town planning: among these, we could mention the integration of configurational values with interactional ones, so as to combine the effects of the articulation of the grid with the results of the actual land use, both concurring in determining the urban phenomena. All the same, on this way, it seems a good step ahead.

References

- Batty M (2001) Exploring isovists fields: space and shape in architectural and urban morphology. *Environment and Planning B: Planning and Design* **28**: 123-150
- Bortoli M, Cutini V (2001) Centralità e uso del suolo urbano. L'analisi configurazionale del centro storico di Volterra. ETS, Pisa
- Cutini V (1999a) Configuration and Movement. A Case Study on the Relation between Movement and the Configuration of the Urban Grid. In: AAVV (eds) CUPUM '99 Computers in Urban Planning and Urban Management on the Edge of the Millennium. Proceedings of the 6th International Conference, Franco Angeli, Milano
- Cutini V (1999b) Urban Space and Pedestrian Movement – A Study on the Configurational Hypothesis. *Cybergeo, Revue Européenne de Géographie* **111**, 26 October 1999
- Cutini V (2000) Configuration and Urban Activities Location. A Decision Making Support Tool. In Proceedings of the 2nd International Conference on Decision Making in Urban and Civil Engineering, Lyon, pp 151-162
- Cutini V (2001a) Centrality and Land Use: Three Case Studies on the Configurational Hypothesis. *Cybergeo, Revue Européenne de Géographie* **188**, 26 mars 2001
- Cutini V (2001b) Configuration and Centrality. Some Evidence from two Italian Case Studies. Proceedings of the Space Syntax 3rd International Symposium, Atlanta (USA) 7-11 May 2001, Alfred Tauban College of Architecture and Urban Planning, University of Michigan, pp 32.1-32.11.
- Cutini V (2003) Lines and squares. Towards a Configurational Approach to the Analysis of the Open Spaces. Proceedings of the Space Syntax 4th International Symposium, University College of London, London, pp 32.1-32.11
- Cutini V, Petri M, Santucci A (2004) From Axial Maps to Mark Point Parameter Analysis: a GIS Implemented Method to Automate Configurational Analysis. Proceedings of the International Conference on Computational Science and Its Applications – ICCSA 2004, Assisi
- Euler L (1736) Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Imperialis Petropolitanae* **8**: 128-140
- Hamer M (1999) Well connected. *New Scientist*, 13.11.1999, pp 29-33.
- Hillier B (1996a) Space is the Machine. Cambridge University Press, Cambridge, UK
- Hillier B (1996b) Cities as movement economies. *Urban Design International* **1**: 41-60
- Hillier B (1999) Why space syntax works, when it looks as it should not. *Environment and Planning B, Planning and Design* **26**: 47-68
- Hillier B, Hanson J (1984) The Social Logic of Space. Cambridge University Press, Cambridge, UK

- Hillier B, Penn A, Hanson J, Grajevski T, Xu J (1993) Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B, Planning and Design* **20**: 67-81
- Lee D (1973) Requiem for large scale models. *Journal of the American Institute of Planners*
- Lynch K (1960) *The image of the city*, MIT, Cambridge, USA
- Montello DR (1991) Spatial orientation and the angularity of urban routes. *Environment and Behaviour* **23**: 47-69
- Sadalla EK, Montello DR (1989) Remembering changes in direction. *Environment and Behaviour* **21**: 346-363
- Turner A (2001a) Angular Analysis. Proceedings of the 3rd Space Syntax Symposium, Atlanta, 7-11 May 2001, Alfred Tauban College of Architecture, University of Michigan
- Turner A (2001b) Depthmap. A program to perform visibility graph analysis. Proceedings of the 3rd Space Syntax Symposium, Atlanta, 7-11 May 2001, Alfred Tauban College of Architecture, University of Michigan
- Turner A, Doxa M, O'Sullivan D, Penn A (2001) From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environment and Planning B: Planning and Design* **28**: 103-121

Validating and Calibrating Integrated Cellular Automata Based Models of Land Use Change

Guy Engelen¹, Roger White²

¹ Centre For Integrated Environmental Studies, Flemish Institute for Technological Research (VITO), Mol, Belgium

² Department of Geography, Memorial University of Newfoundland, St. John's, Canada

Abstract

Realistic high-resolution cellular automata based models of urban and regional systems raise significant problems of calibration and validation. In this chapter we examine first the major philosophical and methodological issues involved in the validation of models that produce as output patterns that are complex but non-deterministic due to stochasticity and bifurcations. Some related problems of map comparison that are significant for both validation and calibration are also examined. Calibration problems are then treated in more detail by means of a case study involving an application of the Environment Explorer model of The Netherlands as well as two semi-automatic calibration techniques that were developed in this context. The calibration tools are shown to be useful if imperfect, and even in some cases to outperform manual calibration.

1 Introduction

Classical science has developed within the framework of theories which make precise and testable predictions. Often these theories are deterministic in nature, as in classical mechanics or, to take a social science example, economics; but even in the case of theories that are fundamentally stochastic, as in the case of thermodynamics, the focus was until recently on applications where the predictions are essentially deterministic. In recent years, however, there has been a flowering of new developments in a number of fields, developments which address problems that could not be handled within the traditional framework, or which address old problems but with a degree of sophistication and verisimilitude that was not traditionally possible. These are the approaches that deal with non-linear or self-organizing systems, also known as complex adaptive systems. The structure of these theories may be relatively simple, but the output in general is quite complex, and often underdetermined, like the systems they are developed to explain; in other words, the output tends to be realistic. By their very nature, these theories

are dynamic, since they represent processes, and they are thus necessarily embodied in simulation models.

In recent years a number of powerful land use models have been developed within the self-organized systems approach. Some of these have progressed beyond purely conceptual and theoretical considerations, and aim at realistically representing geographical systems in terms of the processes modelled (Webster and Wu 2001), the geographical detail attained, and the calibration and validation of the modelling outcomes (Clarke et al. 1997; Straatman et al. 2004). This is certainly the case for cellular automata (CA) based land use models. This trend has come with a relaxation of the standard definition of cellular automata and a consequent increase in the complexity and complication of the models developed (Coullelis 1997). It has led to the development of hybrid CA models constrained in their dynamics by coupled models operating at coarser spatial scales (Batty and Xie 1994; Engelen et al. 1995; White and Engelen 1997) and evolving in a finite non-homogeneous cell space: a bounded cell space consisting of cells with different attribute values representing their physical, environmental, social, economic, infrastructural and institutional characteristics (Clarke et al. 1997; Li and Yeh 2000). These hybrid models are gradually becoming important instruments for the assessment of policies aimed at improved spatial planning and sustainable development (de Nijs et al. 2004) as well as scenario-analysis (White et al. 2004; van Delden et al. 2007). Paradoxically, this transition from the purely academic to the practical sphere of application poses new scientific challenges precisely because of the high quality output. Clearly, neither good science nor practical planning and policy making can be based on tools which produce questionable output; rather the tools must be robust and reliable, based on the best available scientific knowledge and data. This raises to the highest level of importance the issue of the calibration and validation of the models. It also confronts the community of modellers—land use modellers in particular—with the many difficulties associated with the calibration and validation of complex land use models.

The calibration and validation problems are closely related. Both rely on prediction—that is, a comparison of a simulation result (the prediction) with the actual state of the system at the forecast time. Calibration consists of adjusting model parameters so that the model behaviour tracks reality as closely as possible over the calibration period. But a calibration should be tested by running the model beyond the calibration period to some later date for which there is data so that again the model prediction can be compared with reality. This latter step is actually the first step in model validation as such. Thus calibration involves two closely related problems:

1. how to compare the output of the model with reality, and
2. how to change parameter values so as to improve the results of the comparison.

The first of these is the validation problem, but it necessarily arises in calibration as well. It is a problem in part because currently available comparison techniques are inappropriate or inadequate. But more fundamental difficulties arise from the very nature of the models to be validated. Models which produce detailed and complex output like a high resolution land use map need to be validated on

the basis of that detail. And the non-linear models which produce the complexity are typically characterized by bifurcations at various scales, so every run of the model will produce a different output—i.e. a different prediction—even with no change in parameter values. How can the many predictions be compared with the one observed reality? This is an area in which major methodological advances are needed, and this need is generating work in several fields, including geography and the philosophy of science.

2 The Validation Problem

The new non-linear, self-organizing land use models, as well as the theories that they embody, are difficult to fit into the methodological framework of traditional science. Traditional science progresses by hypothesis testing, and ideally the hypothesis is generated by a predictive theory. In this approach, both the hypothesis itself and, more generally, the theory that generates it, are forced to represent the observed world; if they fail at that, they are modified or rejected. In this perspective relationships are assumed to be simple (e.g. linear, logistic, etc.) but obscured by noise in the data. In the new models, however, noise is not error, but an essential feature of the system. Furthermore, the behaviour of the system may be so complex that what looks like noise may actually be dynamics.

2.1 Spatial Bifurcations

Complex land use models necessarily contain a stochastic perturbation. At the micro scale, this produces noise in the predicted land use patterns, noise that appears as minor variations in the land use configurations from one run of the model to the next. However, due to the non-linear nature of the dynamics, the land use patterns generated by the models are also subject to meso- and macro-scale bifurcations, and at the bifurcation points the noise will determine in a random way which path the system will take. In other words, the model will generate a different land use prediction each time it is run, and while many of these will be highly similar, the results will fall into similarity classes, with members of one class being significantly different from members of other classes. If model results are displayed as a series of probability maps cumulating the output of many runs of the model, one map for each land use, then a particular map, say for residential land use, will typically show that for large areas the model produces quasi-deterministic output, with probabilities either very high (>90%) or very low (<10%), but in some patches the probabilities may be intermediate in value, thus revealing a spatial bifurcation. For example, Fig. 1 shows a probability map generated by repeated runs of a CA based land use model applied to Dublin, Ireland. The green areas are regions which the model predicts may or may not experience urban development—in many runs certain parts of the green regions were converted to urban use, but in many other runs, other parts of the green areas were urbanized instead.

In actuality, some of these areas will almost certainly be developed while others will not, but the model cannot determine which will in fact be developed. The bifurcation represented in this example is not very dramatic, so Dublin is in this sense relatively predictable. In general, however, bifurcating systems are only weakly predictable in the sense that the future possibilities open to the system may be predictable, and it may even be possible to predict approximate probabilities for each of these possible futures, but the particular future that will be realized can not be predicted.

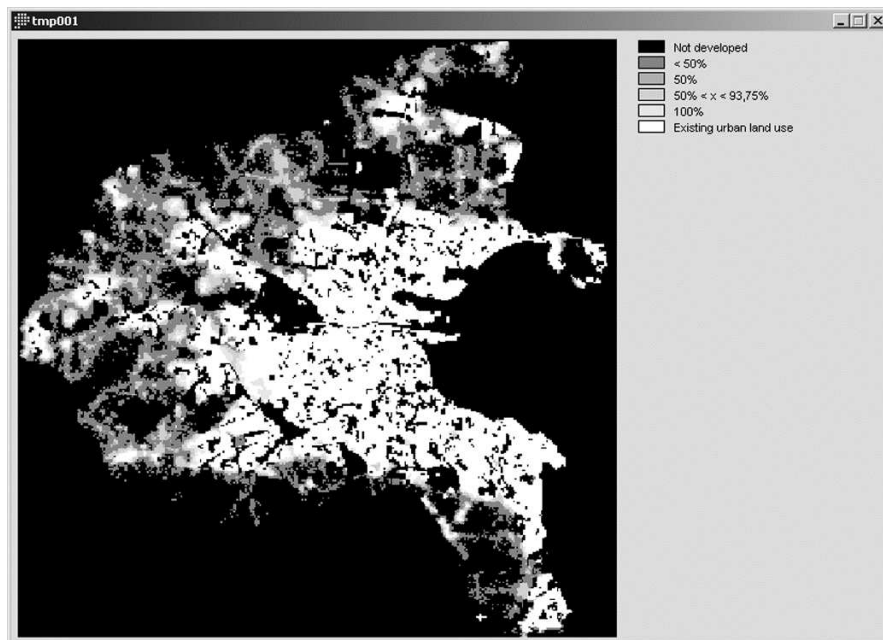


Fig. 1. Probability of urban development 2000-2025 from repeated runs of a stochastic CA based land use model.

This situation, of course, raises problems for both calibration and validation. Brown et al. (2005) has shown that the bifurcation phenomenon undermines attempts to extract CA transition rules from pairs of land use maps (one earlier, one later) using statistical techniques. The statistical approach will mis-calibrate the model because it will produce parameter values that cause the model to yield the observed land use configuration in almost every run, whereas a correct calibration would produce the observed land uses less frequently, and the rest of the time would produce the pattern corresponding to the other fork of the bifurcation.

In the case of validation, if significant residential areas on the actual land use map correspond to areas of very low probability for residential land use on the model-generated map, then it is very likely that the model is either wrong or badly calibrated—in either case, it fails the validation test, so there is no methodological

problem. But in the case where the model is not falsified, how is it to be validated? To the extent that the actual residential areas correspond to regions predicted to have intermediate probabilities of residential land use, then the model is not falsified, but it is only weakly validated. The problem is that for any given application of the model, we can have only one empirical observation to test it against. One solution is to apply the model to a number of different cases (cities, regions). To the extent that the calibrated values of the model parameters are similar in the different applications, and the actual land use map in each case corresponds to areas of reasonably high probabilities as predicted by the model, then the validation becomes stronger. While this approach has not yet been implemented formally, the generic version of the CA land use modelling framework described in Parts 3 and 4 of this chapter has been successfully applied to a number of cities and regions (e.g. Cincinnati, Dublin, Vienna, Prague, Lagos, The Netherlands, and the Italian province of Pordenone), and in each case the calibrated values of the land use transition parameters are, *mutatis mutandis*, relatively similar, thus strengthening confidence that the model does in fact capture something of the essentials of the actual land use change process.

2.2 Multiple Tests

The rich and realistic detail produced by CA based simulation models opens the door to another strategy for dealing with the problem of weak validation: multiple tests involving different aspects of the output. Classical approaches to spatial structure such as the Alonso-Muth (Alonso 1965) model of urban land use or more recent variations such as the new urban economics (Papageorgiou 1976) make clear, testable predictions, e.g. that the various land uses in a city are arranged in concentric circles, or that density declines with distance from the central business district. These predictions are true, as can be established by standard statistical tests. But they are extremely weak. The predictions hold for every city, and consequently the models cannot distinguish one city from another. The CA based models, however, produce detailed land use maps and thus ostensibly make much more powerful predictions in that they not only implicitly predict the Alonso-Muth land use zones and a density gradient, they also predict much more detailed features of the land use pattern. They predict, so to speak, that Dublin should look like Dublin and not like Manchester or Bilbao.

As Karl Popper (1965) pointed out, it is logically impossible ever to validate a theory; the best that can be done is to fail to falsify it. However, the stronger the test that fails to falsify the theory, the more confidence may be placed in the theory. From this perspective, the new land use models are, *a-priori*, more falsifiable, since they predict not just one, but a variety of features of the land use pattern, each of which can be tested, and this is a positive characteristic in terms of the validation issue. Moreover, there is a growing realization among the philosophers of science developing the framework known as evolutionary epistemology (Radnitzky and Bartley 1987) that the methodological belief in rigorous hypothesis testing is misplaced—that such rigorous testing is possible only in relatively trivial

cases. And the dynamic land use simulation models and the emerging theory of geographical systems which they embody are not trivial cases. The implication of evolutionary epistemology is that it is impossible ever to know absolutely that a theory is true. Knowledge evolves by a process of trial and error, and the elimination of positions which prove less satisfactory in favour of those which in practice are more satisfactory. In this view a theory typically is not validated by a single definitive test, but by a series of tests which individually are not conclusive, but together build confidence that the theory has a useful degree of validity.

Following this argument, validation of a land use model can best be approached by applying it in as many situations as possible, and also testing as many independent characteristics of its output as possible. The multiple applications of the CA land use model have already been mentioned. In terms of the independent output characteristics, the map features which can be described and thus be used as the basis for validation tests range from the very specific to the very general. At the extreme of specificity are pixel based tests. At the other extreme are tests involving characteristics of the map as a whole, like the radial dimension and the cluster size frequency distribution. Falling between these extremes are pattern based techniques like polygon matching, feature association, and local fractal measures.

2.3 Local Tests of Map Similarity

Since CA based land use simulation models produce highly detailed maps showing perhaps two dozen land uses at a resolution typically of 100 to 500 metres, depending on the area to which the model is applied, any comparison of the simulated land use with the actual should ideally make full use of this detail. But that is not such a straightforward objective.

The simplest way to compare two maps is pixel by pixel or, using CA terminology, cell by cell. Such a comparison of two land use maps, the simulated versus the actual, for Dublin for the year 2000, is shown in Fig. 2a. But the cell by cell approach is not really appropriate for comparing maps for validation purposes, because it does not distinguish between disagreements due to noise, the small random variations in the precise location of activities on the map, and disagreements that represent differences in pattern or form. Hagen (2003) and Hagen-Zanker et al. (2005a) have attempted to deal with this problem by developing a fuzzy version of the cell-by-cell comparison. In this approach the map comparison is not Boolean: the land use on corresponding cells on the two maps is not recorded as simply the same or different, but rather a level of agreement between the corresponding cells is calculated, with the agreement being higher if the same cell state is found in close proximity on the two maps. This fuzzy comparison of the two Dublin maps is shown in Fig. 2b. Visually, this approach suggests that the level of disagreement between the two maps is less serious than would appear on the basis of the Boolean comparison in Fig. 2a. The yellow areas are regions where the maps are similar but not identical—regions where it might be said that the simulation is qualitatively good but errs in detail.

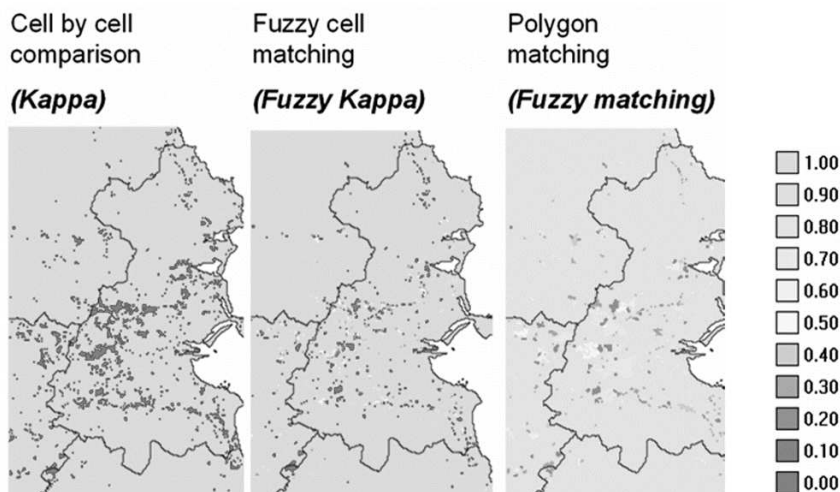


Fig. 2. Results of three map comparison techniques: left (a), strict cell by cell comparison; centre (b), fuzzy cell by cell comparison; right (c), polygon based comparison.

Power et al. (2001) moved away from the cell-by-cell approach to local comparisons by developing a polygon based approach. One map, typically the actual land use map, is taken as the reference, and then, using a fuzzy inference framework, a measure of the degree of overlap of land use polygons on the two maps is calculated for each polygon, with small polygons being weighted less in order to filter out noise from the pattern. The Dublin comparison using this technique is shown in Fig. 2c. Of the three approaches, this is the most effective in filtering spatial noise to reveal areas of real disagreement between two maps. In other words, polygon matching maps show areas where the *pattern* is different on the two maps. This is important because land use patterns express functional aspects of the socio-economic system generating the landscape.

In all three cases a global index value can be calculated to show the degree of similarity of the two maps: Kappa, Fuzzy Kappa, and for the polygon matching technique, the global matching value. In the case of the maps compared in Fig. 2 these values are all quite similar, but in general they can differ substantially. While these global values abstract completely from the spatial configurations shown on the maps, they are useful in automatic calibration routines, since they can be used as a selection criterion for parameter values: keep a new trial value of a parameter if it raises the relevant index value, since that means a better fit between the simulation and the actual map.

But the polygon based approach is only one way of looking at land use patterns. Many others are possible. Almost any morphological characteristic could constitute a basis for comparing two maps. The question is: what pattern qualities are important? Is the irregularity of the edges of land use polygons significant? In an ecological context, this may have great functional significance, and this is one of

the measures included in the FRAGSTAT software (McGarigal et al. 2002). Is cluster shape important? In some calibration of land use models polygons tend to become progressively rounded, so that the 'style' of the simulated map is not realistic, even while the agreement between the simulation and the actual land use map is very good by standard measures such as the Kappa index or the polygon based global matching index. Is a relatively fine scale intermixing or confusion of land uses a significant landscape quality? From the standpoint of the New Urbanism planning perspective it is (Dutton 2000), since it is a land use morphology characteristic that is associated with pedestrian friendly development. Are isolated cells of a particular land use just noise, or are they functionally significant? They may be both, since some such cells act as seeds for larger, growing clusters and thus are critically important for the future development of the land use pattern at the regional scale. (Fig. 3) This phenomenon, in fact, underlies one of the global measures of structure, the cluster size frequency spectrum discussed in the next section.

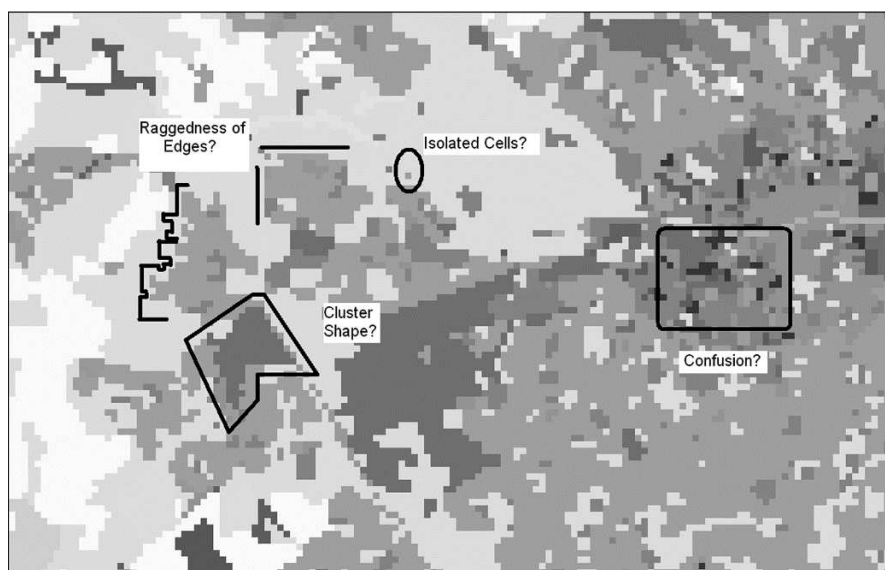


Fig. 3. Examples of possible pattern qualities for use in map comparisons.

In general, there are no explicit criteria for selecting morphological characteristics on which to base map comparisons. Features have been selected on an ad-hoc basis by researchers in a variety of fields to serve immediate needs, and this is perhaps a useful way to proceed. But to the extent that the processes generating a land use pattern are understood, then a useful guide would be to choose to measure those pattern characteristics that are important to the functioning of the system. This pragmatic approach is acceptable for map comparisons used in the calibration stage. However, comparisons based on map metrics are generally felt to be

inadequate for purposes of validation, since map metrics in general are not associated with underlying probability distributions, and hence do not permit a formal test of a hypotheses with an associated confidence level. This is the heart of the formal validation problem: we have no means for accepting, at a given probability level, that two maps, the simulated and the actual, are, except for noise, equivalent—that is, that they are generated by the same process. Furthermore, it seems unlikely that rigorous, probability based tests of detailed map characteristics will ever be available. However, from the perspective of evolutionary epistemology, this is likely to be a false problem.

2.4 Global Measures of Map Similarity

Cities and regions, like most self-organizing complex adaptive systems, typically have fractal structures. In fact there are a number of ways in which land use patterns prove to be fractal in nature: for example, in terms of the area-perimeter relationship (Batty and Longley 1994), in terms of local density gradients as measured by the correlation dimension (Frankhauser 1994), in terms of global density gradients (Frankhauser 1994; White and Engelen 1993), and in terms of the cluster size frequency relationship (White 2006). The classical rank-size relationship of urban geography is also a signature of fractal structure in urban systems. Because the global measures of fractal dimension tend to be robust with respect to spatial bifurcations, they have some value as tools for validation. This value is limited, however, because these measures achieve their immunity from the effects of bifurcations by being only very general, and hence very weak, measures of spatial morphology. In other words, two maps with land use patterns that appear to be completely different may have almost identical fractal dimensions.

The radial dimension and the cluster size frequency dimension in particular are proving to be useful in the calibration and validation of CA based land use models. Both of these are global measures, characterizing the land use pattern as a whole. The radial dimension is primarily useful in connection with models of urban areas, or more generally in connection with any spatial structures that have grown outward from a central site. The radial dimension expresses the rate at which the area occupied by the structure increases as the radius is incremented. Essentially all cities, for example, have an urbanized area in which the built-up area increases with approximately the 1.92 power of the distance from the centre of the city. Beyond this zone is an area in the process of urbanization, in which the built up area increases with approximately the first power of distance from the centre. The radial dimension is particularly useful in calibrating the parameter controlling the level of stochasticity in urban land use models, since it is highly sensitive to the level of stochasticity. It is straightforward to adjust the value of the parameter until the radial dimension of the simulated urban area is the same as that measured for the actual city.

Individual land uses also show the bi-fractal radial structure, but the values of the exponents relating area to radius are different for each land use. For example, the discontinuous sparse residential land use is absent in the centre of Dublin, and

then, once it appears, increases with the 3.72 power of the radial distance from the centre, out to a certain distance, where the rate of increase falls to the 1.4 power; beyond that zone is the rural hinterland (Fig. 4). Other land uses have their own characteristic area-radius relationships, and, the ordering of the different land uses in terms of their inner zone radial dimension, from lowest to highest, is also an ordering in terms of their relative location in the city. In other words, the radial dimensions show that different land uses are in fact, in a statistical sense, arranged in concentric zones around the centre, as predicted by classical land use theory. Since this relative zonation of the land uses always appears in the simulation output from the urban land use models in spite of the fact that the models were developed without any explicit mechanism to produce such a macro-scale zonation, this result in effect contributes to a validation of the model.

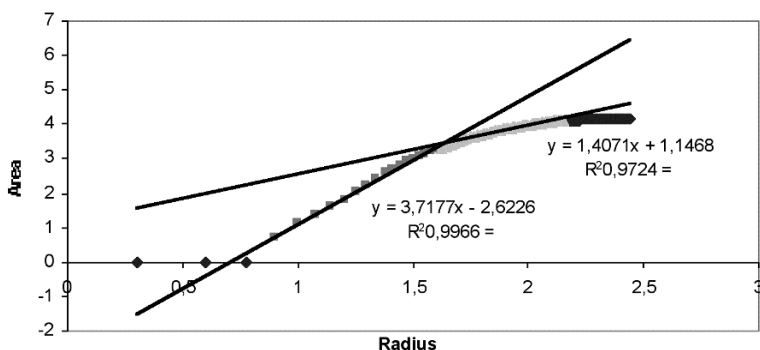


Fig. 4. The radial dimension: area-radius graph for discontinuous sparse residential land use, Dublin, 1988; axes are logs; the slopes of the two regression lines are the radial dimensions of the two zones.

The cluster size frequency dimension expresses the relationship between cluster size and the number of clusters of that size. Typically there are many small clusters, a moderate number of mid-sized clusters, and only a few very large ones. The relationship is log linear and so is characterized by a single exponent (Fig. 5). The cluster size frequency dimension can be calculated for regions as well as cities, since it does not depend on a growth centre for its measurement. Like the radial dimension, this relationship is useful in calibrating the stochasticity parameter, since too little stochasticity in the model system means that seeds for new land use clusters are not generated and the linear relationship is not maintained when the model is run for long periods. And the fact that the model is able to produce a stable relationship amounts to a validation of the model, although a relatively weak validation, since again, there is no explicit mechanism in the model for producing any particular cluster size frequency relationship.

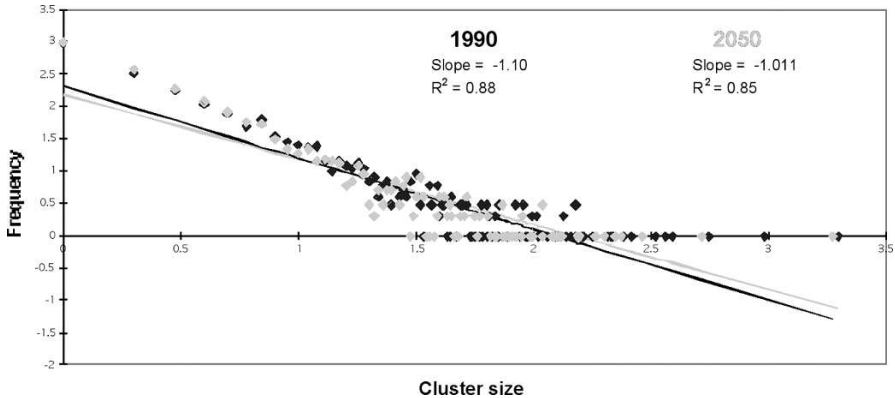


Fig. 5. The cluster size frequency relationship for urban clusters in The Netherlands; 1990 is actual, 2050 is from the land use simulation.

3 A Case Study of Calibration: ENVIRONMENT Explorer

Notwithstanding the many unsolved problems and pitfalls associated with land use models and their calibration and validation discussed in section 2, there is a growing need for these instruments to support planning and policy making. Consequently, in recent years, a growing number of projects are aiming at the development of practical land use modelling tools. One such initiative is Environment Explorer. It started in 1997 and since has generated a sophisticated modelling framework used for the design, analysis, and evaluation of policies relative to the physical environment in the Netherlands in an economic, social and ecological context. In this section we describe the model, and in section 4, the calibration techniques used to make it operational.

The Environment Explorer modelling framework consists essentially of a dynamic land-use-transportation model representing processes operating at three hierarchically embedded geographical levels (Engelen et al. 2003). At each level, a representation can be chosen adapted to the precise needs of the problem studied and the available data in terms of the degree of sophistication and spatial resolution applied: the Global (one spatial entity: typically the Netherlands), the Regional (n administrative entities: provinces, municipalities, or aggregates of the latter) and the Local (N cellular units, ranging from $1/4$ ha to 4km^2 in resolution).

3.1 Global Level

At the Global level, the scale of the entire modelled area, the model runs on exogenously provided time series. It integrates data taken from economic, demographic and environmental growth scenarios. Sophistication in producing these

remains external to the modelling framework proper. The main economic activities are typically condensed into 3 to 10 sectors based on their characteristic spatial behaviour, the availability of data at the three levels of the model, and the end-use requirements. Sectors include: farming, industrial, commercial, services, socio-cultural, and recreational activities. The population is typically represented in two to four residential categories, including: high- and low-density residential. While economic and population sectors are represented in terms of numbers of people, jobs, or economic output, the natural land-use categories are expressed in terms of area occupied. They include, typically: wetlands, forests, shrub vegetation and extensive grasslands.

3.2 Regional Level

At the Regional level, consisting of administrative regions (typically the 40 COROP regions), the national growth figures are a constraint for models catering for the fact that regional inequalities will influence the location and relocation of residents and economic activity and thus drive regional development. A *standard potential based model* is applied (see for example: Wilson 1974; White 1977; 1978; Allen and Sanglier 1979a; 1979b): each region competes with all the other regions for residents and activity in each economic sector on the basis of its geographical position relative to the other regions, its employment level, its population size and the type and quantity of activity already present. In addition to these, and novel in the context of interaction based models, three summarized cellular measures, obtained from the model at the Local level, characterise the regional attractiveness. They are the abundance of good quality land (suitability), the availability for development (zoning status) of that land, and its accessibility relative to transportation or other infrastructure.

A *density sub-model* translates the activity and population numbers per sector into claims for land. The latter are passed on to the model at the Local (cellular) level for their detailed allocation. The principle of supply and demand applies to regulate the densification of the land used as well as its spatial allocation. Alternatively, and in particular for the natural land categories and recreational land, a claim for land is fixed and passed on as a hard constraint thus reflecting the fact that policies determine the amount of land that is to be created or to be preserved per region.

A *transportation sub-model*, in the most sophisticated version a complete four-stage transportation model, linked dynamically in the modelling framework, calculates the changes in the characteristics of the transportation infrastructure, the flows of people and goods travelling over it, and their impact on interregional distances and accessibility.

3.3 Local Level

At the Local level, the detailed allocation of land associated with economic activities, residents and natural land cover in each region, is modelled by means of a cellular automata based land-use model of the kind developed by White, Engelen and Uljee (White and Engelen 1993; Engelen et al. 1995; White et al. 1997). Each region is represented as a regular grid of cells representing parcels of land covering an area in the 1/4 ha to 4 km² range, most typically 25ha. Land use is classified into a (software-technical) maximum of 32 categories, subdivided in 'feature states' (land uses that do not change dynamically), 'function states' (land uses changing dynamically as the result of the Local and the Regional dynamics) and 'vacant states' (land uses changing dynamically due to the Local dynamics only). The function states are chosen with a view to guarantee to the extent possible a one-to-one relationship with the economic and residential categories at the Regional level. For each cell in a vacant or function state, the model calculates at every simulation step the *transition potential* for each possible land use function (function and vacant state). Cells will change to the land-use function for which they have the highest transition potential, until Regional demands for the land-use function are met. In the latter case, they will change to the land use for which they have the second to largest transition potential, and so on. The transition potentials are a proxy for the land rent reflecting the pressures exerted on the land. It accounts for the fact that the presence of complementary or competing activities and desirable or repellent land uses is of great significance for the cell's locational quality and thus for its appeal to a particular land use function. To that effect, the model assesses the quality of the cell's neighbourhood: a circular area with a radius of maximally eight cells. For each land-use function, a set of rules determines the degree to which it is attracted to, or repelled by, each of the other functions present in the neighbourhood. The rules articulate inertia, action at a distance, push and pull forces, and economies of scale, in short, the strength of the interactions as a function of the distance separating the land-use functions within the neighbourhood. In addition, the transition potential comprises characteristics of the cell proper: its physical *suitability*, *zoning status* and *accessibility*. Physical suitability and zoning status are represented in the model by one map per land-use function modelled. Suitability refers to the degree to which a cell is physically fit to support a particular land-use function and the associated economic or residential activity. Zoning status specifies whether a cell can or cannot be taken in by the particular land use during a particular period of time. It is important that suitability and zoning are handled separately in view of analysing policy and planning alternatives. Both suitability and zoning are composite measures, prepared in a GIS on the basis of a series of physical, ecological, environmental maps and master plans and other planning documents respectively. Finally, *accessibility* for each land-use function is calculated relative to the transportation infrastructure. It is an expression of the ease with which an activity can fulfil its needs for transportation and mobility in a particular cell and accounts for the distance of the cell to the nearest link or node on each of the networks, the importance and quality of that link or

node, and the particular needs for transportation of the activity or land-use function.

3.4 Interlinked Dynamics Structuring Space

The linkage between the models at the Global, Regional and Local levels is very intense: the Global figures are imposed as constraints on the Regional models, the Regional models distribute and allocate Global figures to the regions and impose the resulting Regional figures on the cellular models. Finally, the cellular models determine land use at the highest level of detail. The cellular model returns to the Regional model aggregated information on the quality and availability of space for each type of economic or residential activity. It is an input into the spatial interaction calculations at the Regional level and influences the relative attractiveness of the individual regions. Regions running out of space for an activity will lose part of their competitive edge and exert less attraction. This framework constitutes a flexible and powerful instrument for representing non-linear spatial dynamics operating across a range of scales. Such flexibility is essential given its use in practical policy and planning exercises as it enables choosing a spatial resolution and representation of the economic sectors, population categories and environmental drivers as a function of the problems and solutions investigated.

Environment Explorer is used typically for analyses on the mid to long term horizon, meaning some 5 to 30 years into the future. It runs on a yearly time step and generates alongside the changing land use a series of social, economic and environmental indicators, each of which becomes available in the course of the simulation as a time series of maps, both at the Regional and the Local level. It thus is an excellent instrument for scenario-analysis (de Nijs et al. 2004) as well as ex-post and ex-ante assessments of spatial policies and planning options, including transportation policies (Geurs 2003).

Monitoring and analysis of land use change and the associated spatial planning problems is set in a dynamic context. Consequently the models need frequent adaptations as new knowledge and data become available and the urgency of problems and themes requiring attention shifts. However, almost any change in a model or its databases necessitates at the least a partial *recalibration* and *revalidation*. Calibration and validation are prerequisites for any model to be reliably used in practical planning and policy-making. For Environment Explorer, they are both time consuming tasks because of the interlinked nature of the dynamic models used. Since 2005, Environment Explorer has been equipped with facilities supporting calibration and validation. This has been an upgrade much longed-for. The next section will provide more details on the latter.

4 Semi-Automatic Calibration and Validation of Environment Explorer

Calibration and validation of dynamic land use models of the kind discussed is not a trivial problem for the simple reason that they consist of tens of thousands, if not millions of coupled dynamic equations resulting in a potentially extremely rich behaviour. However, in the calibration process, the task is precisely to ensure that the model is able of generating observed and existing spatial patterns. Typically a historic calibration is carried out in which the model is iteratively tested and tuned to generate a spatial behaviour observed in a past period documented in available maps, time series and other data (Hagen-Zanker et al. 2005b). A historic calibration will require a sufficiently long calibration period, typically some 10 years, so that the underlying processes in the system have time to manifest themselves in a representative manner. The quality of the calibrated model is captured in goodness-of-fit measures expressing the level at which the model is able of generating the known state at the end of the calibration interval. Improving the model is an optimization problem solved iteratively. There is no guarantee that such procedure will result in the unique optimal parameter set. The typically large size of the set as well as the strong linkages among the parameters is responsible for this. Calibrating complex models consisting of several sub-models can be simplified by decomposing it into its elementary components and calibrating the latter prior to calibrating the coupled model. While used in the decoupled mode, the sub-models are exogenously fed with data mimicking the outputs of the otherwise linked sub-models. This approach guarantees that the erroneous behaviour of the model is as much as possible detected and corrected in the right sub-model and that goodness-of-fit measures maximally attuned to the specifics of each sub-model are used to that effect. Environment Explorer is thus equipped with routines for calibrating the Regional model, the Local model and the linked Global-Regional-Local model (Engelen et al. 2005). As the Global model consists of a set of consistent time series representing growth scenarios, no calibration as such is required. It suffices to apply the set of time series reflecting the Global trends during the calibration interval.

4.1 Calibration of the Regional Model

As discussed in the previous section, the Regional model consists primarily of an *Activity* sub-model coupled to a *Density* sub-model per sector modelled. All sectors are interlinked intensely. Activity in a region is the result of its attractivity, which in itself is essentially a potential calculation. Activity can be concentrated in a region on more or less land as determined by the density sub-model. The activity sub-model has –depending on the precise sector– 13 to 16 parameters, while the density sub-model has 9 parameters. For a model with as little a 5 sectors this quickly adds up to over 100 parameters requiring calibration. The algorithm applied for the semi-automatic calibration of the regional model implements state of

the art hybrid optimization methods: principles of genetic algorithms, golden section search, and random search (van Loon 2004). It consists of one local optimizer (Golden Section Search) and one global optimizer (Random Search) applied to escape from local minima. The goal of the calibration procedure is to find parameter values that minimize the difference between the actual values of the state variables and the values obtained by running the model. The goodness-of-fit (Eq. 1) is measured by means of a squared sum of relative errors, both for activity X and density W . Relative errors are used because activities and densities in different regions, economic and residential sectors vary greatly in magnitude. Squared rather than absolute errors are applied to emphasize outliers. By means of weighing parameters, the function can be specified to accentuate the final (w_{End}) and/or an intermediate state (w_{Extra}) in the calibration interval, specific sectors (w_K), and parameters of the activity model (w_β) and the density model (w_δ). Thus, the automatic calibration routine can be made to highlight particular states, sectors and sub-models, possibly to home in on one or the other sub-model while performing particular steps during the optimization process.

$$\begin{aligned}
 & w_{End} \cdot \left[\sum_{K=1}^{NrSectors} \sum_{i=1}^{NrRegions} \left(w_K \cdot \left(w_\beta \cdot \left(\frac{EndXref_{Ki} - EndX_{Ki}}{Xinit_{Ki}} \right)^2 + w_\delta \cdot \left(\frac{EndWref_{Ki} - EndW_{Ki}}{Winit_{Ki}} \right)^2 \right) \right) \right] + \\
 & w_{Extra} \cdot \left[\sum_{K=1}^{NrSectors} \sum_{i=1}^{NrRegions} \left(w_K \cdot \left(w_\beta \cdot \left(\frac{ExtraXref_{Ki} - ExtraX_{Ki}}{Xinit_{Ki}} \right)^2 + w_\delta \cdot \left(\frac{ExtraWref_{Ki} - ExtraW_{Ki}}{Winit_{Ki}} \right)^2 \right) \right) \right] \quad (1)
 \end{aligned}$$

With:

$EndX_{Ki}, EndW_{Ki}$

Activity (X) and density (W) of sector K in region i as generated by the model at the end of the calibration interval.

$ExtraX_{Ki}, ExtraW_{Ki}$

Activity (X) and density (W) of sector K in region i as generated by the model at an intermediate state in the calibration interval.

$EndXref_{Ki}, EndWref_{Ki}$

Actual activity (X) and density (W) of sector K in region i at the end of the calibration interval. They are the goal of the calibration.

$ExtraXref_{Ki}, ExtraWref_{Ki}$

Actual activity (X) and density (W) of sector K in region i at an intermediate state in the calibration interval. They are the goal of the calibration.

$Xinit_{Ki}, Winit_{Ki}$

Activity (X) and density (W) of sector K in region i at the beginning of the calibration period. These values are the denominators in the calculation of the percentage deviation in all regions and sectors. The latter is a good measure for comparison, given the differences in size between regions and sectors.

The automatic calibration procedure is primed with values determining the lower an upper bounds of the parameters. Next, it will search for an initial parameter set on a sector by sector and equation by equation basis, hence assuming

that the model is not coupled. This will further result in an ordering of the sectors based on those that require more or less calibration effort: the ones requiring most effort will typically be dealt with first in the sequential procedures. Following the initialization, the main procedure is started. This is essentially an iterative loop which runs until it is halted by the analyst, after a number of preset iterations, or when the goodness of fit is no longer improving. Iterations will typically converge to a set which is 'near to optimal' although there is no absolute guarantee that the optimal set will be found. Each iteration will perform a local search followed by a random search. The local search applies hill climbing by means of the Golden Section Search method to find a local optimum for each parameter of the model. Sectors are dealt with in a so-called 'revisiting mode' meaning that the procedure will begin with tuning one sector's parameters only, next it will extend the parameter set with those of a second sector, next a third sector, and so on until the parameter set is complete. It thus enables to gradually introduce the linked nature of the equations and the dependant nature of the parameters. When incorporating a new sector, the weights given to the sectors already incorporated is changing: the last incorporated sector gets the biggest weight, while the other sectors get an equal but lesser weight. For example after 4 steps in which respectively sectors 1, 2, 3 and 4 have been incorporated, in step 5, sector 5 will be given a weight $w_5 = 3/4$ and $w_1 = w_2 = w_3 = w_4 = (1-w_5)/4 = 1/16$. The procedure will deal with the parameters of the activity and density sub-models sequentially. To that effect, the weights ω_{β} and ω_{δ} will be set equal to 1 or 0.

In its search for optimal parameter values, the local optimiser will carry out in each iteration a search on the complete domain of the parameter as it is set in the initialisation phase of the calibration procedure. To that effect, the range of possible values is cut in $NrTries$ consecutive parts of equal length. For each of these sub-domains the local optimiser searches for a local optimum by means of the Golden Section Search method. The local optimum with the best goodness of fit is considered, provisionally, the global optimum for the parameter. It is used in the remainder of the procedure for further improvement.

The Random Search procedure is incorporated in the calibration procedure in order to escape from local optima rather to find global optima. It is applied after the local optimiser has generated values for the complete parameter set. The method applied is based on the mutation step of genetic algorithms. It starts with determining the two most influential parameters. For each, two new values are selected randomly within the range of possible values: one larger and one smaller. The value with the best goodness of fit is chosen to continue the procedure and begin a new local search. If this new local search results in better goodness of fit, then the new parameter set is withheld in the remainder of the procedure.

4.2 Calibration of the Local, Cellular Model

While the calibration of the Regional model makes use of state of the art, yet generic optimisation procedures, the calibration procedure developed for the Local model is fully attuned to the specifics of the cellular automata model used at this

level. It is a further development of the procedure described in Straatman et al. (2004). It produces a rule set, a series of interaction-values, that generates a model output with the best possible goodness of fit. While the calibration methodology for the Regional model is a fairly well finished product, the calibration procedure for the Local model can still be much improved.

In order to discuss the calibration procedure it is required to return to the specifics of the cellular model and explain the formal expression of the transition potential. The transition potential is a dimensionless value (either positive or negative) representing the likelihood that a cell will change state. As explained, it is calculated for every cell and every function and takes into consideration the neighbourhood quality, as well as cell-specific characteristics: physical suitability, zoning status and accessibility.

The neighbourhood quality (${}^tN_{l,c}$) is calculated using Eq. 2.

$${}^tN_{l,c} = \sum_{d \in D} \left(\sum_{x \in C_d} w_{l,{}^tLU_x,d} \right) \quad (2)$$

With:

D The set of discrete distances in the CA-neighbourhood of cell $c \in C$.

C_d The set of cells at distance $d \in D$, with $C_d \subset C$.

tLU_x The land use of cell $x \in C_d$ at time t .

$w_{l,m,d}$ The interaction-value between cell $c \in C$ with land use l and cell $x \in C_d$ with land use m on distance d in the CA-neighbourhood.

Thus the neighbourhood quality sums the interaction-value between land use l and every other land use that occurs in a cell x which is part of the set of cells C_d at distance d , for every d in the CA-neighbourhood of cell c .

The transition potential (${}^tP_{l,c}$) is calculated for every land use $l \in L$ and cell $c \in C$ by:

If ${}^tN_{l,c} \geq 0$

$${}^tP_{l,c} = \left(w_S \cdot S_{l,c} + w_Z \cdot {}^tZ_{l,c} \right) \cdot {}^tA_{l,c} \cdot {}^tN_{l,c} \cdot {}^tR_{l,c} \quad (3)$$

If ${}^tN_{l,c} < 0$

$${}^tP_{l,c} = \left(w_S + w_Z - \left(w_S \cdot S_{l,c} + w_Z \cdot {}^tZ_{l,c} \right) \cdot {}^tA_{l,c} \right) \cdot \left({}^tN_{l,c} \cdot {}^tR_{l,c} \right) \quad (4)$$

In both cases, the stochastic perturbation factor ${}^tR_{l,c}$ is determined by

$${}^tR_{l,c} = -{}^{10} \log \left({}^trand_{l,c} \right)^a \quad (5)$$

With:

w_s, w_z	weights for respectively physical suitability and zoning status
$S_{l,c}$	Physical suitability
${}^tZ_{l,c}$	Zoning status
${}^tA_{l,c}$	Accessibility
${}^trand_{l,c}$	is drawn from a uniform (0, 1) distribution, with $P({}^trand_{l,c} < x) = x$
a	Stochastic factor

The target of the calibration procedure is to improve the model to the extent that it is capable of generating a land use map that is as similar as possible to the actual land use map at the end of the calibration interval. Rather than tuning the cell and land use function specific terms suitability, zoning status and accessibility, it is the goal to address the generic and dynamic element of the model, which is the neighbourhood quality. For the sake of the calibration, Eqs. 3 and 4 can be simplified to Eq. 6, where $\alpha_{l,c}$ is a multiplication factor representing all the factors in the formula that are constant for a certain land use function (l) and cell (c) during a calibration iteration. Hence, ${}^tN_{l,c}$ is the only factor requiring attention in the calibration procedure.

$${}^tP_{l,c} = \alpha_{l,c} \cdot {}^tN_{l,c} \quad (6)$$

Since the model will assign the land use function to the cell for which its transition potential is highest, until all demands for the function are met, the aim of the calibration routine is to change the neighbourhood quality, and more specifically the constituting interaction-values, to the extent that the n cells having this land use function on the actual land use map are precisely the same n cells with the highest transition potential for that land use function as calculated by the model. The interaction-values are specified as distance functions, described by 5 parameters only as displayed in Fig. 6. The neighbourhood quality for a particular land use function l in a cell c is the sum of the area of the polygons formed by the points (0, 0), (0,a), (1, b), (c,d), and (e,0) of all distance functions of l and all other land use functions $m \in L$ present in the neighbourhood of the cell. Changing the interaction values, means moving the points, this will change the area below the distance curve and hence will change the neighbourhood quality and the transition potential. This is what the procedure does.

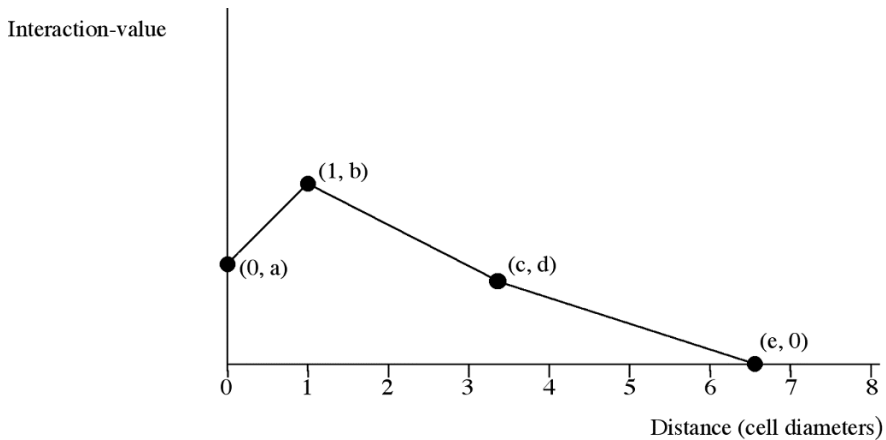


Fig. 6. Typical distance function of the interaction-values between a pair of land uses in the neighbourhood with a radius of 8 cells. Each distance function is defined by 4 points and 5 parameters (a, b, c, d and e).

To this effect, the procedure deals with three types of land uses. It will try to remove land use l_{calc} , as calculated by the model, from places where in the actual land use map l_{act} is found. It does so by changing the influence of a third land use l_{infl} on either l_{act} or l_{calc} by modifying the respective interaction-values. There are a number of smaller steps to be taken to achieve this, which are described at length in Engelen et al. (2005). Briefly stated, the procedure will first deal with the land use l which is closest to being correctly positioned on the calculated map. The reasoning behind this choice is that the interaction-values for this land use require the least changes, and, that improving the positioning of this land use may have as a secondary effect a better positioning of other land uses too. Next the procedure will find out which other land uses are typically overrepresented in the neighbourhood of cells with the selected land use l_{act} . It does so by comparing the composition of their neighbourhood with those of all cells with another land use. From this comparison it will select the pairs (l_{act}, l_{infl}) that are overrepresented most, hence considered most influential for attracting the selected land use to the neighbourhood. Next, the interaction values are tweaked by changing the parameters a, b, c, d and e one by one with an amount guaranteeing that the transition potential rises above the threshold so that the cell changes state to l_{act} , or that the transition potential for l_{calc} drops below this threshold. This results in $2 \times 5 = 10$ new sets to be tried out. Again, the procedure will start with the cell which transition potential is closest to the threshold. It is repeated land use pair after land use pair over and over again. Every time interaction-values have been changed, the goodness of fit is calculated to ascertain that the changes made are in fact improving the results globally and not locally only. If none of the modifications in the 5 parameters of a land use pair results in a better goodness of fit, then the pair is put on a revolving taboo list so that is not being selected for a while.

The goodness of fit between the actual map and the calculated map is measured by means of the Fuzzy Kappa (Hagen 2003). Better than the Kappa Statistic (see among others Monserud and Leemans 1992), the Fuzzy Kappa enables taking into consideration the importance of deviations between the actual and the calculated map. As discussed in section 2, it allows for small misplacements of land uses –e.g. not in the proper cell, but in its immediate vicinity–, and similarities of the land use classes –e.g. not ‘high density residential’, but the very similar ‘medium density residential’–.

One problem with the calibration routine described is that it could produce near to perfect interaction-values for a model going in one step from the initial to the final state of the calibration interval. In reality the calibration interval will cover many more years, typically some 10 years. The changes made to the interaction-values are therefore corrected to reflect an annual change. However, given the non-linear behaviour of the model as well as the discrete nature of the land use changes, this is nothing but a poor approximation. As discussed in Straatman et al. (2004) a more ideal, yet in practice unrealistic, situation would be one in which the described routine could be provided with land use maps on a year by year basis, so that interaction-values could be modified in function of the yearly time step. Another problem with the methodology described is that it involves a deterministic procedure to calibrate a model with an important stochastic component, represented by the stochastic perturbation term $'R_{t,c}$. This can be dealt with by applying the procedure in an iterative loop which changes the seed of the random number generator for drawing $'rand_{t,c}$ in Eq. 6. Consecutive iterations will adjust the interaction-values to result eventually in a set which is more robust to the perturbations. It goes without saying that this is a solution requiring a lot of time. It is to be replaced by better alternatives when available.

Even if the cellular automata model is primed with interaction-values that are systematically set equal to zero, the methodology described will generally find interaction-values with a fair goodness of fit. However, while a good model can be defined on the basis of three or four distance functions per land use model, the procedure will typically generate a near to complete set of distance functions, many of which have a very limited role in improving the goodness of fit and some of which can not be given a logical explanation in terms of the structuring processes that they are supposed to encapsulate. That is why the modeller has to intervene in the automatic calibration to remove the interaction-values constituting the latter distance functions. With the reduced set, the routine is started again to fine-tune the remaining interaction-values. Most often, this will improve the overall results, in terms of both technical and theoretically underpinned interaction-values, as the model is forced to concentrate on the important distance functions.

With both the Regional and Local models calibrated in isolation, the dynamic coupling between both is re-established and the calibration of the coupled model begins. This involves testing the goodness of fit of both the Regional and the Local models while coupled. If the goodness of fit of either decays, then, the calibration of the Regional and Local models is started all over. To that effect, the time series, used to mimic the behaviour of the other sub-model and fed in exogenously, are recomputed with the coupled model.

5 Results - The Pilot Case: ENVIRONMENT Explorer for the Netherlands

For Environment Explorer four complete datasets are available for both the Regional and the Local model: 1989, 1993, 1996 and 2000. This enables to calibrate the model over the period 1989 – 1996 with 1993 as the intermediate year. Further to use the 2000 data set for validation purposes. Even though this is seemingly an ideal situation for a calibration exercise, there are many problems with the quality and compatibility of the available data. For example, the dubious quality of the land use map for 2000 is demonstrated in testing the likeliness of the state transitions observed in the period 1996 – 2000 on the basis of the 1996 and 2000 land use maps. It turns out that as much as 25% of the state changes are unlikely to have occurred in reality. They involve transitions from land into water and vice versa, from urban into natural land use, etc. Unfortunately, this situation is not unique to Environment Explorer, rather is more the rule than the exception at this moment. Land use modellers can only insist on getting better data for their models but are not responsible for generating them.

The results of the calibration of Environment Explorer are at first sight promising as is displayed in Table 1. At the Regional level, the activity sub-model allocates 96.1% of all activities (both residential and economic) in the right region, and, the density sub-model is reasonably correctly generating the amount of land taken in per region: on average a mismatch of 3.3 cells per sector modelled. The Local model generates land use maps that have a high similarity with the actual map of 2000 as reflected in the very high Fuzzy Kappa value of 0.94 (1.00 represents a perfect match; -1 is a complete mismatch). For the validation period (1996 – 2000) these figures are generally lower, respectively 94.8% of the activities, a mismatch of 7.7 cells and a land use map similarity of 0.91.

Table 1. Main results of the calibration (1989 - 1996) and validation (1996 – 2000) of Environment Explorer at the Local and Regional levels

	Local level		Regional level		Regional level	
	Fuzzy Kappa [-]		Activity [% growth]		Area [Nbr. of cells]	
	EE	RCM	EE	CS	EE	CS
1989-1996	0.94	0.92	3.9	5.2	3.3	5.7
1996-2000	0.91	0.92	5.2	3.9	7.7	6.4

In order to put this calibration/validation in perspective, and to make better founded statements about the quality and performance of Environment Explorer, the results are compared with those of simpler models, called Naïve Predictors. The performance of the Regional model is tested against that of a Constant Share model (CS), a model that allocates Global growth to each region proportionally to its initial size, and, the performance of the Local model is tested against a Random

Constraint Match model (RCM), a model that adds or removes land uses to the extent that the observed numbers are met, yet positions the modifications on the map randomly. Clearly, the two naïve predictors are simpler models in that they lack the capacity to cluster activities according to spatial interactions principles and specialization. Yet, Table 1 reveals that Environment Explorer barely outperforms the naïve predictors in the calibration period, and, that in the validation period the opposite is true.

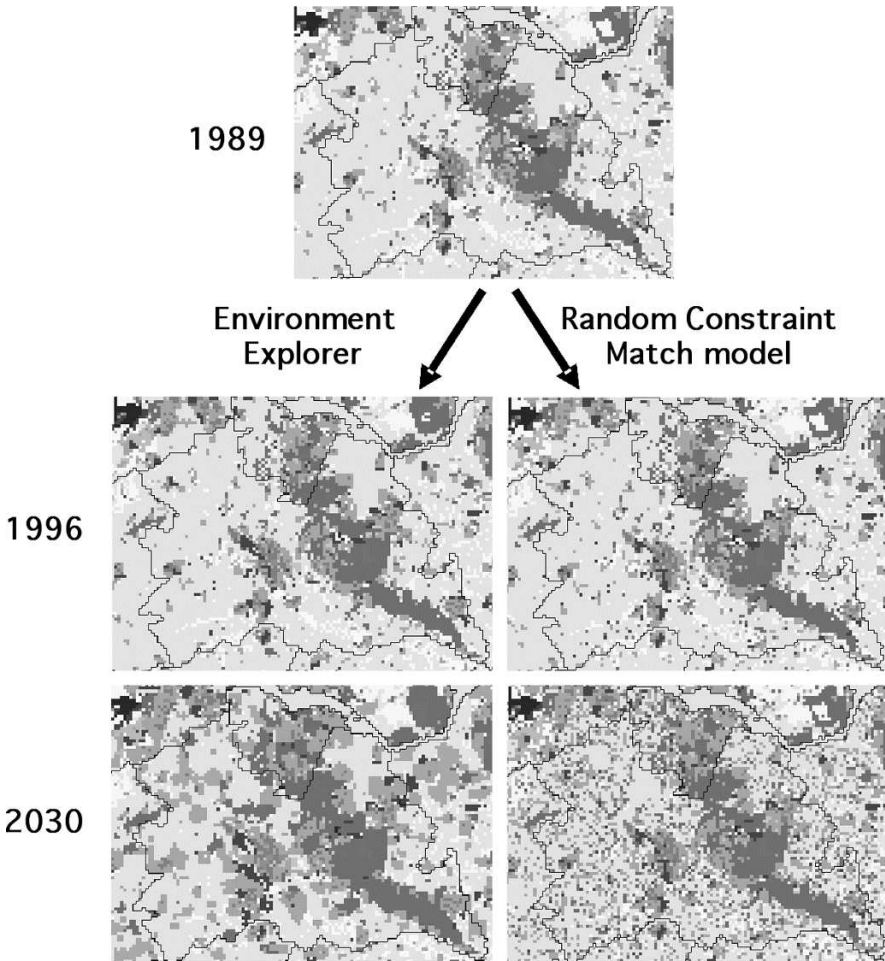


Fig. 7. The effect of time on the results generated by Environment Explorer and the Random Constraint Match model.

Clearly some explanation for this seemingly disappointing performance of Environment Explorer can be found in the dubious quality of the 2000 land use map,

although it is fair to assume that both Environment Explorer and the RCM model are suffering equally of the deficiencies in the 2000 land use map. More of an explanation is to be expected from the short calibration and validation periods applied. It is to wonder to what extent the main structuring processes have been able to manifest themselves in the seven-year span of the calibration period, hence, to what extent they have been detected in the calibration procedure and have lead to the fine-tuning of the main structuring parameters. Clearly in the given circumstances, simple models assuming more of the same, are apparently better in forecasting the future in the short run. For longer periods the availability of a model which is more sophisticated in structuring space may be more of a necessity as Fig. 7 demonstrates. The composition, size and distribution of the clusters generated by Environment Explorer are far more realistic than those of the RCM model.

6 Discussion

Calibration and validation, however inappropriate we can tackle them currently when dealing with self-organizing land use models, are tasks requiring vast amounts of human resources and time. Moreover, models have a tendency to change as the modeller finds new ways to improve the model formulation, new data become available, or, the problems to solve with the model change. The example discussed in this paper, Environment Explorer is not so much a model rather a modelling framework, designed to enable rapid reconfigurations of a core model into versions best adapted to the problem studied. Hence, it runs very frequently in one if not all the problems named: new problems to address, reconfigured models to tackle them, and new data to work with. Consequently, in the 7 or 8 years of its existence Environment Explorer has been subject to much calibration and validation effort and the costs and loss of valuable time associated with it. The design and development of automatic calibration and validation routines was an obvious route to engage on. Currently, the system is equipped with automated calibration procedures in a prototype version. None could be called 'fully automated' since they still require a substantial user involvement. However, when used in semi-automatic mode, the methods speed up the calibration of both the Regional and the Local model considerably: a good parameter set is found in a matter of hours, typically one or two days, rather than one or two person-months of hard and tedious work by a specialist. Moreover, when started from an initial parameter set delivered by a specialist or taken from another application, the methods substantially outperform the specialist in the quality of the parameters and rule sets generated.

7 Conclusions

The problems arising in the calibration of the Environment Explorer are not specific to that model. As dynamic models become more complex, in order to be more realistic—and hence more useful for planning, policy, and other practical applications—they all face similar difficulties. These arise in part from data problems—errors in the available data sets, the limited number of dates for which data are available, and frequently, the short interval on which a calibration must be based. But fundamentally the difficulties are rooted in the methodological problems involved in trying to infer process, here represented by parameter values, from pattern, or even from a short temporal series of patterns. These are problems that may not have an exact solution, but it is reasonable to expect that continued efforts to develop better calibration techniques will eventually result in methods that give satisfactory results. The techniques described in this chapter are, after all, only first attempts, yet they are already useful. Further development will undoubtedly improve their performance, and other approaches, not yet conceived, are likely to be much more powerful.

In part, the difficulties with calibration techniques arise from the map comparison problems discussed earlier in this chapter. An inappropriate map comparison technique may well rate a map as worse when in fact it is better. This can be seen in a comparison of the Fuzzy Kappa statistics in Table 1 with the maps for 2030 in Fig. 7: the Fuzzy Kappa values for 2000 suggest that the random constraint model is better than the Environment Explorer, but the maps for 2030 show clearly that the random constraint model is producing an unrealistic landscape, while the land use pattern generated by Environment Explorer looks quite reasonable. Since systematic calibration techniques must depend on map comparison, improvements in their capabilities depend on development of more powerful and appropriate map comparison techniques. It is likely too, that ultimately good calibrations will depend on the use of a suite of map comparison techniques capturing a variety of morphological characteristics, rather than just one.

Beyond calibration, there remains the validation problem. It is almost certain that this cannot be ‘solved’ by development of the sort of rigorous hypothesis testing techniques based in probability theory that many feel are necessary for modelling to become respectably scientific. But as indicated earlier in this chapter, as physics, biology, geography, and other disciplines increasingly adopt simulation modelling and fractal geometry in order to be able to deal with the phenomena of self-organizing complex adaptive systems, ideas of the nature of science and its methodology are co-evolving with the new approaches. So while on the one hand there will continue to be technical advances that will permit a more thorough empirical evaluation of complex models, on the other, the validation problem itself, as a methodological and philosophical issue, will come to be seen in a new and less threatening light. This is the promise of evolutionary epistemology.

References

- Allen P, Sanglier M (1979a) A Dynamical Model of Growth in a Central Place System. *Geographical Analysis* **11**: 256-272
- Allen P, Sanglier M (1979b) A Dynamic Model of a central Place System-II. *Geographical Analysis* **13**: 149-164
- Alonso W (1965) *Location and Land Use*. Harvard University Press, Cambridge, Mass
- Batty M, Longley P (1994) *Fractal Cities*. Academic Press, London
- Batty M, Xie Y (1994) From cells to cities. *Environment and Planning B* **21**: 31-48
- Brown DG, Page S, Riolo R, Zellner M, Rand W (2005) Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science* **19**: 153-174
- Clarke KC, Hoppen S, Gaydos I (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B* **24**: 247-261
- Couclelis H (1997) From cellular automata to urban models: new principles for model development and implementation. *Environment and Planning B* **24**: 165-174
- de Nijs TCM, de Niet R, Crommentuijn L (2004) Constructing land-use maps of the Netherlands in 2030. *Journal of Environmental Management* **72**: 35-42
- Dutton J (2000) *New American Urbanism*. Skira, Milan
- Engelen G, White R, Uljee I, Drazan P (1995) Using Cellular Automata for Integrated Modelling of Socio-Environmental Systems. *Environmental Monitoring and Assessment* **34**: 203-214
- Engelen G, White R, de Nijs T (2003) Environment Explorer: Spatial Support System for the integrated Assessment of Socio-Economic and Environmental Policies in the Netherlands. *Integrated Assessment*, **4**: 97-105
- Engelen G, Hagen-Zanker A, de Nijs ACM, Maas A, van Loon J, Straatman B, White R, van der Meulen M, Hurkens J (2005) Kalibratie en Validatie van de LeefOmgevings-Verkenner. Report 2550016006/2005, Netherlands Environmental Assessment Agency, Bilthoven, The Netherlands
- Frankhauser P (1994) *La Fractalité des Structures urbaines*. Editions Economica, Paris
- Geurs K, Hagen A, Hoen A, van Wee B (2003) Ex post evaluation of Dutch spatial planning and infrastructure policies. Paper presented at the European Transport Conference, October, Strasbourg, France
- Hagen A (2003) Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science* **17**: 3: 235-249
- Hagen-Zanker A, Straatman B, Uljee I (2005a) Further developments of a fuzzy set map comparison approach. *International Journal of Geographical Information Science* **19**: 7: 769-785
- Hagen-Zanker A, van Loon J, Maas A, Straatman B, de Nijs ACM, Engelen G (2005b) Measuring Performance of Land use Models. An Evaluation Framework for the Calibration and Validation of Integrated Land Use Models Featuring Cellular Automata. Proceedings of the 14th European Colloquium on Theoretical and Quantitative Geography, Tomar, Portugal, September
- Li X, Yeh AGO (2000) Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science* **14**: 2: 131-152

- McGarigal K, Cushman SA, Neel MC, Ene E (2002) FRAGSTATS: spatial pattern analysis program for categorical maps. Computer software program produced at the University of Massachusetts, Amherst. www.umass.edu/landeco/research/fragstats/fragstats.html
- Monserud RA, Leemans R (1992) Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling* **62**: 275-293
- Papageorgiou G (ed) (1976) *Mathematical Land Use Theory*. Lexington Books, Lexington, Mass
- Popper K (1965) *The Logic of Scientific Discovery*. Harper and Row, New York (first published 1934)
- Power C, Simms A, White R (2001) Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science* **15**: 77-100.
- Radnitzky G, Bartley WW (eds) (1987) *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*. Open Court Publishing, La Salle, Ill
- Straatman B, Engelen G, White R (2004) Towards an Automatic Calibration Procedure for Constrained Cellular Automata. *Computers Environment and Urban Systems* **28**: 149-170
- van Delden H, Luja P, Engelen G (2007) Integration of Multi-Scale Dynamic Spatial Models of Socio-Economic and Physical Processes for River Basin Management, *Environmental Modelling & Software* **22**: 223-238
- van Loon J (2004) Towards model-specific automated calibration procedures for the Leef OmgevingsVerkenner model. Unpublished Master's Thesis, Maastricht University, Maastricht
- Webster C, Wu F (2001) Coase, Spatial Pricing and Self-organising Cities. *Urban Studies* **39**: 11: 2037-2054
- White R (1977) Dynamic Central Place Theory: Results of a Simulation Approach. *Geographical Analysis* **9**: 227-243
- White R (1978) The simulation of Central Place Dynamics: Two Sector Systems and the Rank-Size Distribution. *Geographical Analysis* **10**: 201-208
- White R (2006) Pattern based map comparisons. *Journal of Geographical Systems* **8**: 145-164
- White R, Engelen G (1993) Cellular Automata and Fractal Urban Form: a Cellular Modelling Approach to the Evolution of Urban Land use Patterns. *Environment and Planning A* **25**: 1175-1199
- White R, Engelen G, Uljee I (1997) The use of Constrained Cellular Automata for high-resolution modelling of urban land use dynamics. *Environment and Planning B* **24**: 323-343
- White R, Engelen G (1997) Cellular Automata as the Basis of Integrated Dynamic Regional Modelling. *Environment and Planning B* **24**: 235-246
- White R, Straatman B, Engelen G (2004) Planning Scenario Visualization and Assessment: A Cellular Automata Based Integrated Spatial Decision Support System. In: Goodchild MF, Janelle D (eds) *Spatially Integrated Social Science*. Oxford University Press, New York, USA, pp 420-442
- Wilson AG (1974) *Urban and Regional Models in Geography and Planning*. John Wiley and Sons, London

Fractal Geometry for Measuring and Modelling Urban Patterns

Pierre Frankhauser

Laboratoire Théoriser et Modéliser pour Aménager (ThéMA), UMR CNRS 6049, Université de Franche-Comté, Besançon, France

Abstract

Urban growth generates nowadays patterns, which look rather irregular. Planning policy regrets the lack of compactness and density of these agglomerations, but controlling urban sprawl turns out to be difficult. Obviously a new type of spatial organisation emerges, which is rather the result of a self-organisation process to which a high number of social agents contribute. In the present contribution we focus on the use of fractal geometry which turned out to be a powerful instrument for describing the morphology of these patterns.

After an introduction about the context of research, fractal models are presented, which serve as reference models for better understanding the spatial organisation of settlement patterns. Then the methodology for measuring their morphology by means of fractal parameters is explained. Moreover different peculiar topics are considered like a specific approach of urban boundaries. Then an overview is given over results obtained for a couple of agglomerations in different European countries. The interpretation of these results allows establishing links between urban planning policy and pattern morphology. Applying the idea of self-organisation leads to introducing a fractal order parameter for studying the emergent fractal order in urban patterns. The presentation of these quantitative results will be completed by some reflections about how planning concepts based on fractal geometry may help to manage more efficiently urban sprawl.

1 Some Features of Urban Patterns

Since the industrial revolution in the 19th century the way of life has considerably changed. Before, the most important part of the population lived from activities directly related to agriculture, whereas nowadays only a small part of the active population lives from such activities. These changes affected crucially the settlement dynamics as well as their spatial structure. In former times few decision maker like sovereigns or municipalities controlled urban development. The necessity of protection against attacks incited them to choose compact regular plans for these towns like squares or circles. In the 19th century the necessity of defence disappeared and moreover new actors, the industrial investors, influenced urban

growth. Hence the connection to transportation networks became a crucial key-factor for localisation of new build-up zones. Before efficient transport systems existed, the size of the towns remained still constrained, but the introduction of railways and later on of tramways and suburban railway systems gave the opportunity for living at one place and working at another one. The result was a more tentacular growth close to the public transportation network axes. Since motorisation this constraint disappeared as the road network irrigates space in a more and more uniform way. Hence larger and larger areas were affected by that what we call nowadays urban sprawl.

Town planners often deplore that the patterns generated by these dynamics look rather “chaotic” and resemble more ink splashes than regular geometric forms like circles or squares. However despite many attempts to limit urban sprawl, this phenomenon remains a challenge for urban planning (Frank 1987; Rémy 1994; Fouchier 1995).

This may be explained by the fact that a great number of public and private actors with quite diverging interests contribute to the settlement dynamics, like property developers, enterprises, planners, politicians or different kinds of pressure groups. Moreover it seems that the desire to move out of very dense urban areas persists and corresponds to a real social demand. Hence policies trying to increase density in the outskirts or to restrict sprawl are usually contested and local decision-makers may even be willing to give in to market pressure.

Thus we may interpret urban patterns as macrostructures which are generated by a great number of individual decisions or decisions of groups acting on different scales (individual scale, the scale of the administrative units ...). The decisions are linked in a complex way, since actors are usually not completely free in choosing their localisation. Hence urbanisation should be considered nowadays as a *self-organisation process* where great number of individuals or groups interact and generate in this way a macrostructure.

The fact that the patterns generated by urban sprawl show peculiar features and that their aspect is perhaps the expression of a certain actual life-style incites different research groups to study this phenomenon in more details. We just remind some of these research projects to which we refer later on. In the late 1980's and the 1990's basic work was done in the frame of the *Sonderforschungsbereich 230* at Stuttgart (cf. e.g. Humpert et al. 1991; Humpert et al. 2002; Frankhauser 1988, 1994; Schweitzer 1998). More recently T. Sievert (1997) introduced the notion of the “*Zwischenstadt*”. In Great Britain different authors worked on this topic (Batty and Longley 1986, 1994). In France an equivalent notion, that of the “*ville émergente*” has been introduced and in the frame of a research program different topics related to this phenomenon have been analysed (Dubois-Taine and Chalas 1997; Dubois-Taine 2002). More recently a *Cost10*-project focussed on this topic (Dubois-Taine 2004).

In the present contribution we will focus on the morphology of these emerging settlement patterns. In particular we are interested at the question to what extend these patterns are organised according to a certain order principles, despite of there irregular form.

One of the main reasons why these patterns look “amorphous” is the fact that they are a patchwork of structural elements belonging to a great variety of scales reaching from that of buildings to the scale of metropolitan areas. Let us consider as example the metropolitan area of Stuttgart. Even in the rather coarse-grained cartographic representation of the build-up area in Fig. 1 (left) it is possible to discover details of very different size: when following the borderline of an arbitrarily chosen settlement cluster, there appear small bays, which alternate with larger ones in an irregular rhythm. This feature is not peculiar to the particular size of a town, it can be found in small as well as large ones. Moreover the metropolitan areas are made up of many clusters of quite different size, spatially distributed in a rather non-uniform way: along valleys or transportation networks we often observe ribbons of built-up areas, whereas badly accessible zones are sparsely populated. The same characteristics may be found on the micro-scale of towns: in Fig. 1 (right) we have made a zoom on the northern urban fringe of the agglomeration of Stuttgart where industrial areas, historical town centres and recent detached housing areas are mixed. The buildings form clusters of different size and the empty space separating clusters vary within a large range.

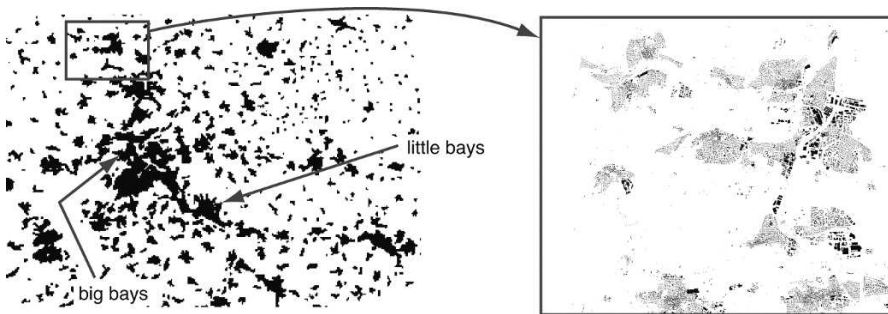


Fig. 1. Left: the metropolitan area of Stuttgart in a simplified cartographic representation; Right: a detailed GIS-data base for outskirts.

For describing the distribution of build-up area usually traditional measures are used based on densities. However density measures the *mean occupation* of space. Hence density is constant when the constitutive elements of a structure, in our case the buildings, are distributed uniformly in space what does not seem to be the case for urban patterns. Moreover density does not really give information about spatial distribution. Figure 2 shows two patterns where 64 squares of same size have been distributed in different ways within the same square: density will be the same, whereas the first pattern is a fractal-like structure, contrarily to the second one.

Similar ambiguities occur when searching reliable criteria for defining urban boundary. When looking at the simplified cartographic representation of Fig. 1 (left) we may expect that urban boundary can be identified without difficulties. However this is no longer the case if we consider the real situation represented on Fig. 1 (right).

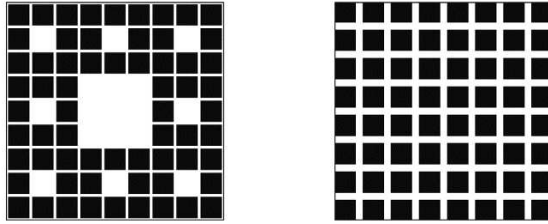


Fig. 2. 64 squares distributed in two different manners: the left distribution follows fractal logic, the right one is uniform.

The urban pattern consists of a great number of detached buildings or groups of buildings and the distances between these buildings vary considerably. This is particularly the case for the fringes of urbanized areas where recent detached housing estates and traditional rural settlement patterns are mixed as shows Fig. 3. Administrative definition recur however to the distances between neighbouring buildings to delimitate urban areas: e.g. in France a settlement is considered as separated from the neighbouring one if the distances between neighbouring buildings are greater than 200m. In other countries similar rather questionable criteria are used. Hence the only physical limits left in urban areas, in a strict sense, are the walls of the buildings. However we will see that by using other geometric approaches we may find criteria for delimiting urbanized areas.

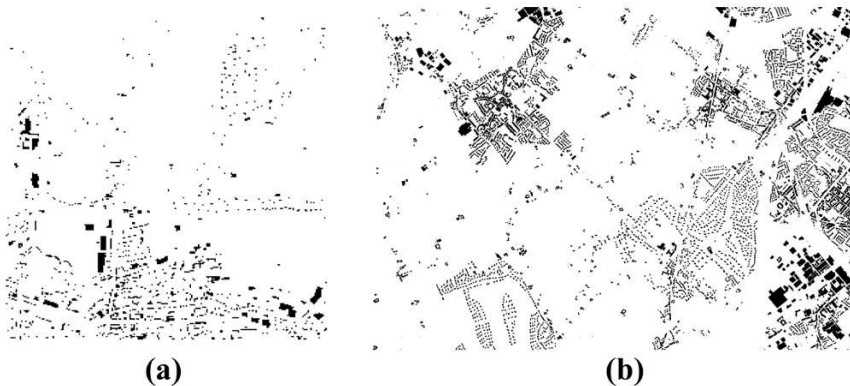


Fig. 3. Left: the northern urban fringe of the city of Besançon (East of France): diffuse strips of houses follow the road network, where ancient farm houses and new houses are mixed. Right: the western part of the urban area of Lille (North of France): different kinds of urban pattern are mixed.

Obviously one difficulty for apprehending the spatial organisation of urban patterns results from the fact that their constitutive elements, the buildings, are distributed in a rather non-uniform way, forming clusters at various scales. Despite their complexity, it is nevertheless possible to discover regularities in these

patterns that seem paradox. If we measure, for instance in cartographic representations like that of figure 1a, the boundary length of urban clusters and their surface, we observe that both the border length and the surface tend to be *proportional*. This clearly contradicts the usual geometric assumption that the surface should be proportional to the square of the perimeter length. Another observation is interesting too. Let us assume we determine in a simplified cartographic representation of metropolitan areas like that of figure 1a the surface of the different clusters. In a second step we introduce surface size classes and we determine the number of clusters belonging to these classes. In many cases, e.g. for Stuttgart, Berlin or Munich, we observe that the number of clusters is linked to their size according to a strong hierarchical principle: there exist few big clusters and an increasing number of smaller and smaller ones. The observed hierarchy follows a hyperbolic distribution law, well-known in urban geography as Pareto-Zipf distribution (Frankhauser 1992; Schweitzer and Steinbrick 1998; Schweitzer and Steinbrick 2002).

Such observations led researchers to turn to a completely different approach for describing the morphology of such patterns, namely one based on fractal geometry. The rationale for using this approach lies in its inherent properties since fractal geometry allows constructing spatial reference models where elements are distributed in a completely non-uniform way forming clusters at different scales. It is then possible to illustrate several types of spatial pattern organisation, which resemble peculiar aspects of urban patterns like fragmentation or complex morphology of boundaries. These structures may look irregular; the spatial distribution of the constitutive elements follows nevertheless a strong distribution law which may be characterized by a unique value. Thus, if urban patterns really show the particular features of fractal objects, we may conclude that despite their highly irregular aspect, they follow a well-defined spatial organisation principle, which can be characterised by quantitative factors. The usual notion of “regularity” or “irregularity” then becomes meaningless. Hence applying fractal geometry helps to find coherent interpretations of the mentioned phenomena and paradoxes.

Basic work on fractal investigations of urban patterns has been done since the 1980's, in particular by M. Batty and P. Longley (1986, 1994) as well as by ourselves (Frankhauser 1988, 1994) and R. White and G. Engelen (1993). More recent publications have deepened the methodological aspects and have confirmed the interest of using this approach (e.g. Batty and Kim 1992; Batty and Xie 1996; Frankhauser 1998; Frankhauser and Genre-Grandpierre 1998; Frankhauser 2000; Benguigui et al. 2000, Shen 2002; Schweitzer and Steinbrick 2002, De Keersmaecker et al. 2003; Frankhauser 2003).

We emphasize that the particular interest of using a geometrical approach is that we are able to develop new spatial reference models which play a similar role as do circles, squares etc. in Euclidean geometry and which allow to illustrate crucial features of urban patterns. Hence we start by introducing such models. Then follows an overview over fractal measures and a discussion of the information they transcribe about urban morphology. This helps to interpret the results obtained from the investigations of the urban patterns and to link observation to the models. Finally, preliminary reflections about how such knowledge could serve

planners in their actual work are presented, in particular with regard to controlling sprawl.

2 Fractal Models for Urban Patterns

Traditional geometry gives information about construction rules which define the typical size of the object: e.g. a circle is defined by the rule that it includes all points situated at a given distance from a fixed point. For fractals the most striking feature is the property of *self-similarity* or *scaling*. Even if a fractal is composed of constitutive elements which may be Euclidean objects like line segments, circles, squares etc., the crucial information is given by the rule which tells us how these elements reappear on a finer scale. This rule is given by the so-called *generator* (Mandelbrot 1983). Thus the *passage from one scale to the next one* is the operation which defines fractal properties. In the present context we will be interested in finding construction rules which allow describing fundamental morphological properties of urban patterns.

The construction starts by defining an initially given Euclidian figure, called *initiator*, but the geometric form of the initiator does not affect fractal properties. The *generator* transforms now this initial figure. In our models we use simple self-similarity rules, defining the following procedure:

1. the size of the initiator, e.g. a square, is reduced by a certain factor r ,
2. N of these little replicates are generated,
3. these N replicates are placed in a defined way, respecting the following restrictions:
 - they must be placed within space occupied by the initial figure, i.e. they must be subsets of the initiator
 - they are not allowed to intersect

We call further on the generated replicates “*elements*” of the structure. In a next step the defined procedure is applied to each of the elements generated in the first step. Fractal geometry stipulates that this procedure is reiterated up to infinity. However when referring to real world situation, we will usually stop this iteration at a certain step n . Mandelbrot called such objects *prefractals*. E.g. if the size of the initial figure corresponds to the area occupied by a town, we will stop iteration as soon as the size of the elements is comparable to that of a building, since we are not interested to model details of buildings.

We first consider the situation we observe on the scale of town sections: the urban patterns consist of detached houses or blocs of houses which are surrounded by streets. Thus we consider a modelling concept which generates detached elements. This is the case for the *Fournier dust*. The way how such a fractal is constricted is shown on Fig 4.

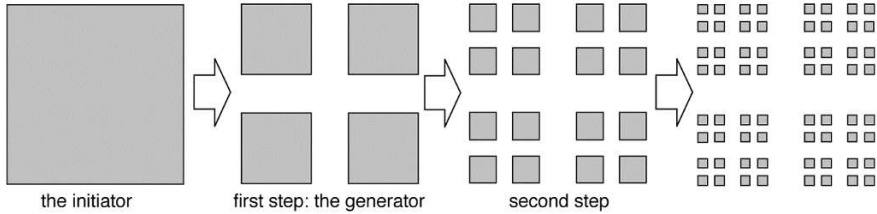


Fig. 4. The iterative construction of a Fournier dust. We observe that the elements in the generator are disconnected

In the example of Fig. 4 we chose a square as initiator. The generator consists of $N = 4$ smaller replicates, the base length of which is $r = 2/5$ of the initial square. Due to iteration the elements are not distributed uniformly on the given area; they form clusters, which are separated by empty lanes of different size: there exists a few number of large lanes and an increasing number of smaller and smaller ones. The relation between the number of lanes and their size follows a strong hierarchical law, which corresponds to a hyperbolic distribution. We verify, too, that the number of elements increases according to a geometrical series in course of iteration. In Fig. 5 we have compared the 2nd iteration step of a Fournier dust with an urban pattern.

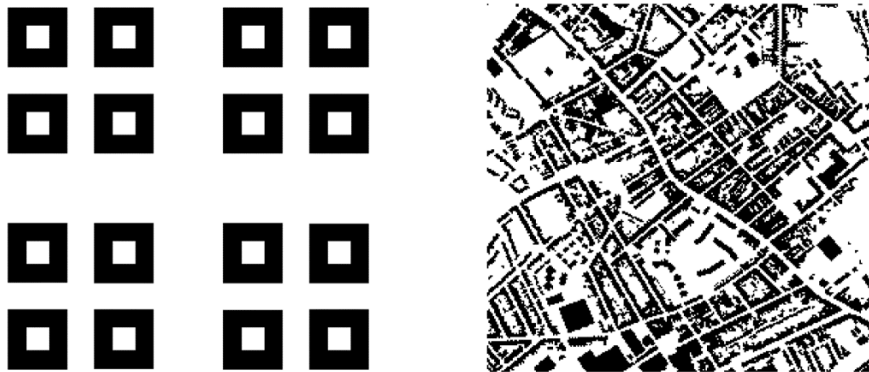


Fig. 5. Comparing a generator of a Fournier dust with urban patterns: the urban pattern consists of terraced houses forming blocks of houses, surrounded by street. This main feature reappears in the fractal.

Since the form of the initiator does not affect fractal properties, which are solely conditioned by the generator, we chose for the upper example a black square, from which we have cut off a square-like central part. This allows symbolizing a block of houses with an inner court-yard.

Figure 6 makes evident the different role of initiator and generator. We compare two Fournier dusts for which the 1st step seems to be identical. In figure (a)

the initial figure is a square surrounded by a white lane, whereas for figure (b) it is simply a black square. Since the reduction factors are different for both generators, the results obtained for further steps are different. In case (b) the lanes separating the elements follow a hierarchy, what is not the case for (a).

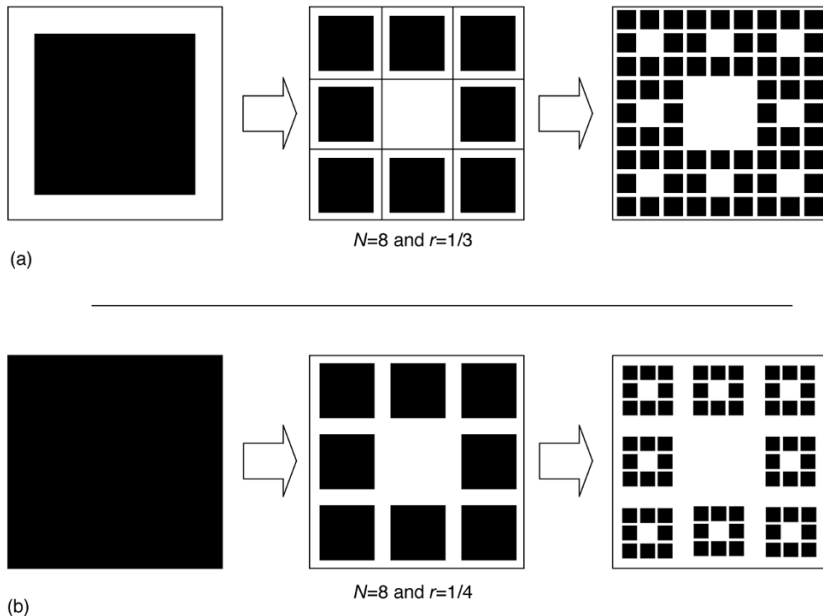


Fig. 6. Two generators with the same numbers of elements N but different reduction factor: the difference becomes obvious for the next iteration step. To ensure good readability, we have suppressed the surrounding lines delimiting the elements in the below figure of (a).

By considering the two different types of cartographic representation for Stuttgart, we have observed that both at the micro-scale of town sections and the macro-scale of the agglomeration we observe clusters of elements separated by lanes of different size. However at the scale of the agglomeration we identified another feature, peculiar for this level of observation: since it is possible to identify clusters, we got aware of the tortuous form of their boundaries. Such characteristics do not exist for Fournier dusts, since their elements are completely detached. However a similar type of fractal, the so-called *Sierpinski carpets*, allows modelling such forms. The only – but crucial – difference with respect to Fournier dusts is the position of the elements in the generator which are now touching each other (Fig. 7a). This holds by the way also for the generator of Fig. 6a which is in fact a *Sierpinski carpet*. Hence a *Sierpinski carpet* consists of one unique cluster all over the iteration steps. For the *Sierpinski carpet* of Fig. 7a we observe that the borderline becomes more and more tortuous in course of iteration.

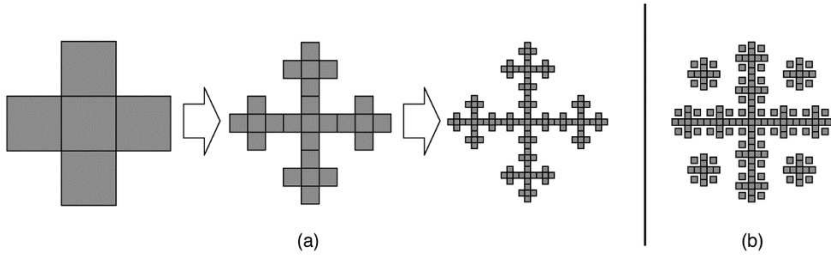


Fig. 7. (a) generating a Sierpinski carpet : all elements touch their neighbours; (b): a hybrid Sierpinski carpet (2^{nd} iteration step)

However, as pointed out, agglomerations consist usually of a hierarchical system of clusters of different size. Such structures may be obtained by combining both the logics of Sierpinski carpets and Fournier dusts; we call such fractals “*hybrid Sierpinski carpets*”. An example is given in Fig. 7b, where the generator consists of $N^{(\text{clust})} = 9$ attached elements which form a cross and $N^{(\text{island})} = 4$ isolated elements. At each iteration step an increasing number of smaller elements are generated and at the same time the borderlines become more and more tortuous since each of the clusters is itself a Sierpinski carpet.

It is possible, too, to generate such fractals by means of a so-called “*outer mapping procedure*”: replicates of the initiator are put around it thus defining the generator. This procedure is repeated in further steps (Fig. 8).

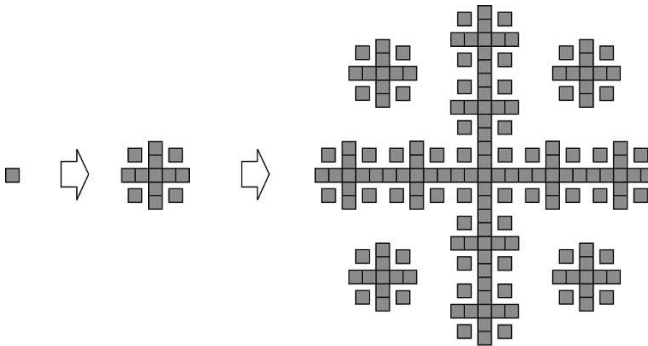


Fig. 8. Generating a hybrid Sierpinski carpet by an outer mapping procedure

Here the steps could be interpreted as different stages of urbanisation: a town, compact at the beginning, develops outskirts along main transportation network axes as well as in the hinterland of eventually pre-existing small settlements. The number of clusters is linked to their size according to a strong hierarchical law, a Pareto-Zipf law: there exist more and more small clusters. This corresponds to the previously enounced empirical observations for metropolitan areas.

We finally introduce another model to which we recur sometimes as reference, too, the *teragon*. This fractal is generated in a different way. The initial figure is in this case a line segment which is reduced by a factor r . A number N of these elements is put together and forms a tortuous line. This generator is again applied to all segments. Hence a more and more complex curve is generated (Fig.9a). If four such line segments are connected and form at the beginning a square, in course of iteration a closed tortuous curve is obtained the length of which tends to infinity (Fig. 9b). The morphology of the borderline reminds the form of urban clusters resembling to ink-splashes.

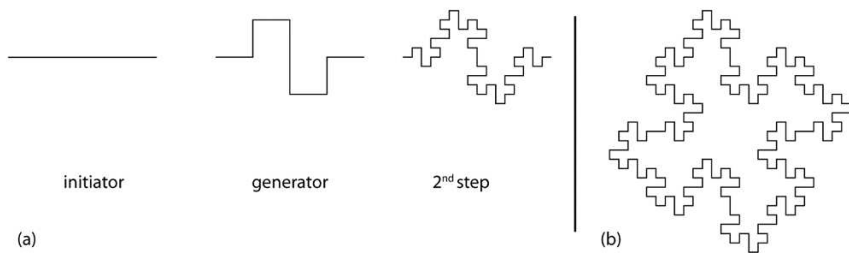


Fig. 9. (a) the construction of a segment of a teragon. (b) The teragon is obtained by assembling four segments

Hence we may conclude that some models seem more to suit to certain scales or topics than to other ones. Fractal analysis of urban patterns should inform us if there exist really critical scales where the spatial organisation changes.

In order to illustrate the main features of our models we generated rather symmetric patterns. Real world patterns do of course not look as symmetrically but it is possible to conceive less symmetric generators as shows Fig. 10. Moreover in course of iteration the position of the elements may be changed, they must only lie within the elements generated in the previous step. Thus we obtain patterns which resemble more empirically observed ones. Such fractals are called “random fractals”.

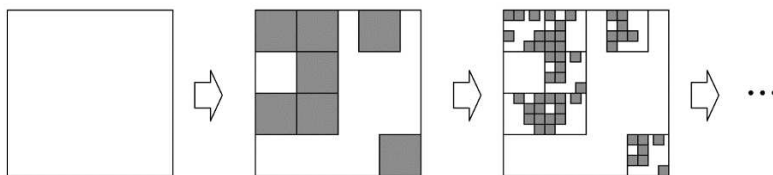


Fig. 10. Generating a random fractal: the positions of the elements are less symmetric. Moreover, for each element generated, the positions of the smaller elements generated at the next step, are again modified. However already generated lacunas remain strictly non-occupied.

3 Measuring Fractal Behaviour of Urban Patterns

3.1 The Notion of Fractal Dimension

We will now see that the concept of iterative mapping, which served to construct our models, is closely linked to the paradox properties of fractals as well as to the parameters which help to describe them. If we compute the length of the borderline of a teragon or of a Sierpinski carpet, we get aware that it tends to infinity in course of iteration even if these objects occupy only a finite part of space. Thus such an object is neither a linear object with dimension one nor a surface with dimension two. On the contrary, the surface of a Sierpinski carpet tends to zero whereas that of teragon remains constant (Mandelbrot 1983).

Such curious properties incited mathematicians to introduce the concept of *fractal dimension* which introduces a generalized concept of dimension. It is not the goal of this contribution to discuss the details of this concept which is largely discussed in other publications (cf. e.g. Mandelbrot 1983, in the context of urban pattern analysis cf. Batty 1994; Frankhauser 1994, 1998). For constructed fractals, like those of our models, the fractal dimension D allows linking the number of elements N_n generated at iteration step n to their size l_n , i.e. in our examples the base length of squares by the relation

$$N_n \cdot (l_n)^D = \text{const} \quad \Rightarrow \quad N_n = \text{const} \cdot (l_n)^{-D} \quad (1)$$

We remind that the numbers N_n and the lengths l_n follow both geometrical series:

$$N_n = N^n \quad \text{and} \quad l_n = l_0 \cdot r^n \quad (2)$$

where l_0 is the length of the initial figure. It is possible to insert these relations into Eq. 1 what yields

$$D = -\log N / \log r \quad (3)$$

This shows that D is a constant parameter all over iteration since for constructed fractals D depends obviously only from both the constant parameters N and r of the generator; we emphasize that the position of the elements in the generator does not affect the D -value. We should be aware that normal geometrical objects, like line sections, squares etc. may also be generated by iteration. For such object relation Eq. 3 yields the usually topological dimension values i.e. $D = 1$ for a line section and $D = 2$ for the surface of a square. *Thus usual geometry is a limit case of fractal geometry.*

Let us focus on the meaning of the fractal dimension in particular for urban patterns. If we consider the spatial distribution of the build-up area, the dimension measures the *degree of concentration* across scale or, more precisely, the *relative decrease of build-up area with increasing distance from any site where mass is concentrated*. Hence, the *uniformly* the elements are distributed in a fractal structure, the closer the dimension will be to *two*, and vice versa, if the mass were concentrated in one point, D_s would be *zero*. E.g. in the generator of Fig. 6a the ele-

ments are more uniformly distributed than in that of Fig. 6b. Consequently the dimension value is higher for the generator (a) ($D = 1.89$) than for (b) ($D = 1.5$).

For hybrid Sierpinski carpets it is possible to define two dimensions: the one describes the distribution of all elements in space. E. g. for Fig. 7b $N^{(tot)} = 13$ and $r = 1/5$ and we obtain $D_s^{(tot)} = 1.59$. However we may separately compute the dimension of the cross-like central clusters, composed of $N^{(clust)} = 9$ elements. This dimension amounts to $D_s^{(clust)} = 1.37$ and characterizes the multi-scale dendritic aspect of this cluster, but also of all the smaller ones generated in course of iteration.

It is also possible to calculate the fractal dimension of the boundary of fractals. For Sierpinski carpets the dimension of boundaries is equal to that of the surface. This may express the fact that in the course of iteration boundary and surface tend to the same limit set. We argued earlier that observation has shown that for metropolitan areas the length of the outline of the clusters is proportional to the built-up surface; from a fractal point of view this would mean that both the edge and the surface have the same fractal dimension, like a Sierpinski carpet.

The situation remains the same for hybrid Sierpinski carpets. Hence $D_s^{(tot)} = D_b^{(tot)}$ describes the spatial distribution of the boundaries of the clusters, or more concretely, the *fragmentation* of the boundaries whereas $D_s^{(clust)} = D_b^{(clust)}$ measures their *tortuousness*.

For the teragon the situation is different, the border has the dimension $D_b = 1.5$, but as pointed out, the inner surface has the dimension *two*. The fractal dimension of boundaries measures *the relative lengthening of a boundary* between two chosen points when comparing it with a straight line relying these points. Often different features of the boundary may be distinguished. For the fractal (a) in Fig. 6¹, the boundary is smooth, its dimension is $D_b^{(extern)} = 1$ whereas for the whole boundary including all lacunas the dimension is $D_b = D_{surf} = 1.89$.

3.2 The Methodology of Fractal Analysis for Urban Patterns

For urban patterns linking the D -value to some iteration parameters is no longer possible as towns are not the result of iterative mapping. Mathematicians have developed a couple of methods which serve to verify to what extent an observed pattern may be considered at least as *prefractal* and, if this is the case, how the dimension may be determined. Since we don't tackle further on with constructed fractals, we prefer sometimes speaking of *scaling exponent* instead of fractal dimension. In the present paper we will confine ourselves to presenting two of these methods which has served to obtain the results presented later on.

For the following we assume having at disposal digitized numeric data of urban patterns. The minimal resolution is then defined by the size of the pixels. We consider binary images, where black pixels correspond to build-up space and we test whether the spatial distribution of black pixels correspond to a fractal law.

¹ Let us remind that the elements of the generator are the squares delimited by lines, and *not* the interior black squares!

A basic method which has been widely used for urban pattern analysis is the so-called *radial analysis* (Batty 1994; Frankhauser 1994). A counting point is chosen, e.g. the centre of town, and then the number $N(\varepsilon)$ of black pixels is counted which lies within a square² of base length ε centred on the counting point. The procedure is repeated for other values of ε . For fractals a similar relation between to Eq. 1 is obtained:

$$N(\varepsilon) = \text{const} \cdot \varepsilon^{-D} \quad (4)$$

Radial analysis provides specific information about spatial organisation around the counting point; hence we speak of local analysis. Contrarily the other method we used recently, the *correlation analysis*, informs us about the fractal mean behaviour within a chosen area. The procedure is the same as before, but counting is realized for all black pixels lying within the chosen area instead for only one counting point. Then for each value ε , the mean number of black pixels situated at distance ε is computed. For these mean numbers the same type of fractal law is obtained as Eq. 4.

For simple fractal structures, the results obtained by both the methods tend to be the same. However we may not expect that real world patterns are structured according a strong fractal law. They may show *multifractal* behaviour: then the fractal behaviour may vary from one site to the neighbouring one. The scaling exponent obtained by radial analysis depends then from the counting point and corresponds rather to the *Lipschitz-Hölder exponent*, whereas the correlation dimension, which is a so-called “second-order” dimension, tests the mean presence of pairs of black pixels lying within a distance range ε .

For urban patterns we may expect to find different types of spatial organisation according to the context of urbanisation: housing estates are often constructed following a certain planning concept and should morphologically be homogeneous. However neighbouring town sections may be issued from different periods or planning concepts and thus be different. Hence we could expect that fractal behaviour may change at certain scales, i.e. when passing from the scale of town sections to that of the agglomeration. Moreover we may of course not exclude that the spatial organisation of a town section is not fractal at all.

Beyond these phenomena local deviations from fractal law may be observed: it is e.g. known that presence of huge empty zones may affect the fractal law. Such deviations could be taken into account by introducing a supplementary parameter, the prefactor a , in the fractal law (c.f. e.g. Gouyet 1992):

$$N(\varepsilon) = a \cdot \varepsilon^{-D} \quad (5)$$

² In principle it is possible to choose any shape for the environment, such as a circle, a hexagon, etc. However, since pixels are square-like, the choice of a square helps to avoid rounding errors.

3.3 The Curve of Scaling Behaviour

In order to identify critical scales where fractal behaviour changes a certain type of representing empirical data turned out to be useful. For illustrating the method we recur to Eq. 5. In order to be able to use differentiation algorithms we consider continuous distance values $\varepsilon_i \equiv \varepsilon$. We take the logarithm of Eq. 5 and we put $x \equiv \log \varepsilon$, $y \equiv \log N(\varepsilon)$ what yields the relation

$$\log N(\varepsilon) = \log a - D \cdot \log \varepsilon \Rightarrow y = \log a - D \cdot x$$

and we obtain

$$\frac{d \log N(\varepsilon)}{d \log \varepsilon} \equiv \frac{dy}{dx} = -D$$

The slope of the $y(x)$ -curve is constant if the parameters a , D are constant. However if we assume that the prefactor a as well as the fractal dimension D depend on the distance parameter ε , we would obtain (Frankhauser 1998):

$$\begin{aligned} \frac{d \log N(\varepsilon)}{d \log \varepsilon} &\equiv \alpha(\varepsilon) = \frac{d \log a(\varepsilon)}{d \log \varepsilon} - \frac{dD(\varepsilon)}{d \log \varepsilon} \log \varepsilon - D(\varepsilon) \\ \Rightarrow \frac{dy}{dx} &= \frac{d \log a(\varepsilon)}{dx} - \frac{dD(\varepsilon)}{dx} x - D(\varepsilon) \end{aligned}$$

Hence two additional terms contribute now to the derivative which we call now $\alpha(\varepsilon)$:

- the first one refers to the local deviations of fractal laws like large lacunas,
- the second one describes the changes of fractal behaviour

Due to these terms the α -values may exceed the upper limit value of $D = 2$. Empirical observations as well as theoretical reflections tell us that changes of the fractal dimension correspond to a progressive change of the mean behaviour of the $y(x)$ -curve which affects a larger range of distances (Frankhauser 1998). This incites us to affect local fluctuations around a dominant mean behaviour of the curve to the first term, but slowly drifting of this mean behaviour to real changes of the dimension D^3 .

This shows that for empirical analysis the changes of $\alpha(\varepsilon)$ is rather instructive. Thus different authors used the representation of the sequence of α -values as instrument for evaluating whether the dimension remains constant over scale or whether changes occur (Batty 1992; Frankhauser 1998). We called the function $\alpha(\varepsilon)$ *curve of scaling behaviour*. We used it mainly when recurring to radial analysis and particularly when introducing a fractal order parameter. Moreover in

³ Comparing the $\alpha(\varepsilon)$ -curves for radial analysis and for correlation analysis shows that for the latter one local fluctuations are quasi absent and only structural changes in D remain. This is due to the fact that correlation analysis refers to mean values, and thus local deviations compensate.

recent investigations we recurred to this curve for selecting zones for fractal analysis: for chosen points radial analyses had been realized and only if for sufficient large zone fractal behaviour is constant we determined the correlation dimension as shown on Fig. 11.

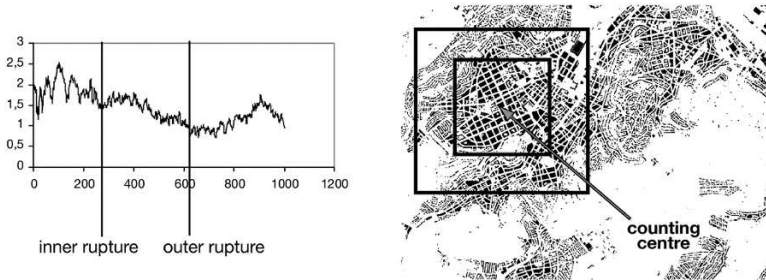


Fig. 11. At left the curve of scaling behaviour obtained for a radial analysis referring to a counting centre localized in the city centre of Stuttgart. Two distances for which ruptures may be identified are indicated on the map at right. The correlation analysis has been realized only for the inner zone, for which fractal behaviour seems rather constant.

3.4 Methodology for Extracting Urban Boundaries

As pointed out before, urban boundaries do not really exist. However fractal geometry allows developing a coherent approach of urban boundaries. According to the logic presented on Fig. 5 we interpret the urban pattern as a Fournier dust for which the iteration has been stopped at a certain step n . It is then possible to associate to this Fournier dust another “virtual” Fournier dust. His generator consists of the same number and the same type of elements as those of the first one, e.g. squares. The elements of the second dust are centred on those of the first one. The only difference is the size of the elements: for the second dust we chose a base length which allows just filling up the empty lanes of size λ_1 separating the elements. Hence we can define a borderline which surrounds all elements. Figure 12c illustrates the procedure for the simple Fournier dust of the Figs. 12a and 12b: the base length of the initiator of the first dust is l_0 , that of the elements of the generator l_1 , whereas the base length of the covering squares is l_1' . This covering procedure can be repeated for further steps as shown for the second one in Fig. 12d: now the lanes of width λ_2 are filled up. This corresponds indeed to generating *another Fournier dust*, for which the parameters N and r , and thus D , are the same as for the first one. At each step the smallest existing lanes are filled up by the second dust and borderlines can be defined which surround blocks of elements of the first Fournier dust (Fig. 12d). Assuming that iteration has been realized up to step n , what corresponds to the urban pattern, we may imagine going backwards to steps $n-1, n-2 \dots$ up to the generator and thus find larger and larger clusters until finally one unique cluster subsists what corresponds to the initial figure and which

we interpret as the “envelope” of our town (for details cf. Frankhauser and Tannier 2005).

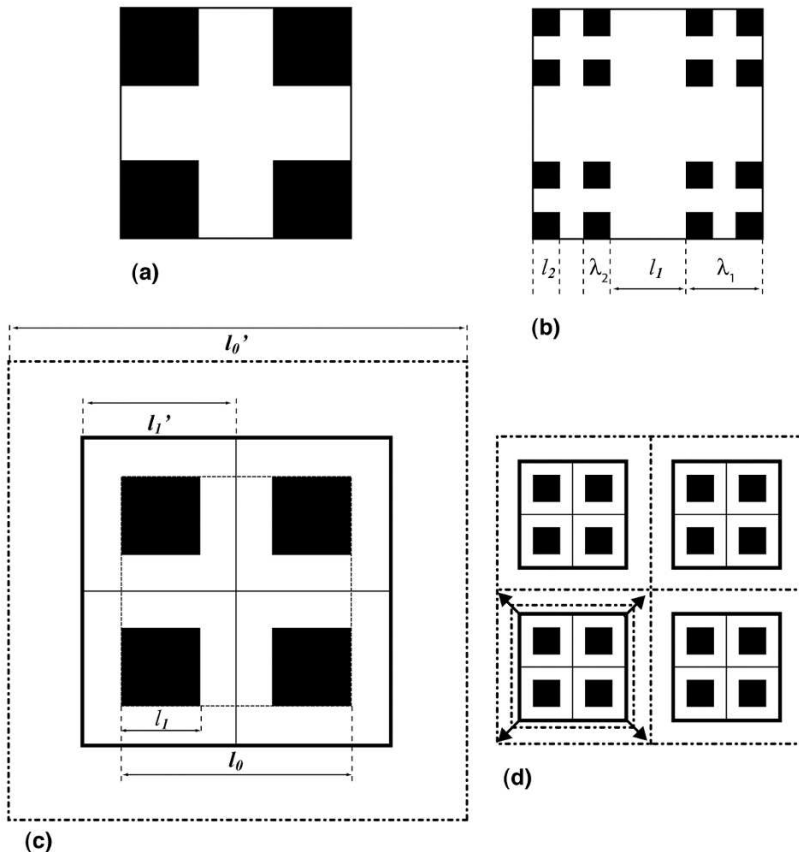


Fig. 12. (a) and (b): The two first steps for generating the Fournier dust of Fig. 1. The widths of the lanes generated are indicated. (c) and (d): The covering of the Fournier dust of Fig. 1 by another Fournier dust. This second dust covers at each step just the lanes generated at the same iteration step for the first dust. The respective lengths of the elements are indicated, where the prime refers to lengths of the second dust.

For real world patterns the covering procedure is not really applicable but it can be replaced by a similar method: we may dilate progressively the urban pattern by surrounding the black surfaces at each step by lanes of width of one pixel. Hence larger and larger clusters appear. It is then possible to extract at each step their boundaries⁴. We may then analyse their tortuousness by means of fractal analysis.

⁴ From a mathematical point of view both the methods are not really identical: the covering corresponds rather to the notion of Hausdorff-Besikovich dimension, whereas dilation may be associated to the notion of the Minkowski-Bouligand dimension.

This allows to test whether the scaling behaviour changes when bigger and bigger clusters appear. An alternative method is the measure the cumulated length of all the boundaries for each step and to verify its variation. this analysis turns out to be of particular interest: if the patterns is mainly structured according to the logic of a Fournier dust the length of the boundaries could increase in course if dilation, but on the contrary, if the pattern resembles more a Sierpinski carpet the length should decrease. Figure 13 shows an example for which the structure changes its character when going beyond the first dilation step. Hence for small distances which refer to the first dilation step, the pattern looks rather like a Fournier dust; all buildings are separated. When having filled small streets and courtyards clusters occur and the spatial arrangement of these cluster dominate further steps. The quasi linear decrease of the boundary length in the bi-logarithmic representation is a hint that the boundary follows a hyperbolic distribution according to fractal geometry.

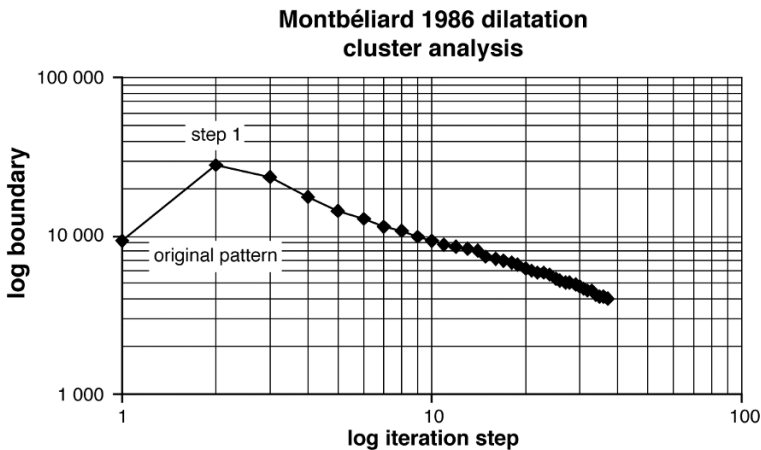


Fig. 13. The changes of boundary length of whole the clusters for different iteration steps for the conurbation of Montbéliard. After the first dilation step the tendency is inverted (cf. text).

4 Some Empirical Results

4.1 The Sample of Towns Presented

As mentioned before, different authors have presented results of fractal analyses of urban patterns. In the present paper we will refer mainly on results we obtained recently in the frame of a research project financed by the French Ministry of

Planning⁵. The main goal of this project was to explore to what extent planning policy and more generally the historical context in which a town has evolved influence the pattern morphology. About fifteen European metropolitan areas located in France, Germany, Belgium, Switzerland and Italy⁶ have been considered for which cartographic data on the scale of buildings were available. Even if the data quality was not strictly the same in all cases, it was possible to compare the results obtained for the different agglomerations (Frankhauser 2003, 2004).

In the present overview only the reliable results are taken into account. The examples presented relate particularly to the agglomerations of Lille, Lyon, Stuttgart, and Montbéliard. This information will be complemented by focussing on some aspects of agglomerations which show particular features like Saarbrücken, Strasbourg and Basel, all located near national borders, the new town of Cergy-Pontoise, some peripheral zones of Brussels and Helsinki, and an Italian case, Bergamo.

Without going into details we give just some information about the particularities of the main urban areas as far as this is of interest for discussing the obtained results. The urbanisation of the metropolitan area of *Lille* was over a long time dominated by heavy industry. The pattern is a rather disparate patchwork of zones of different morphology; transient areas fill space between the hard core of the conurbation and the rural hinterland. More recently, the “new town” *Villeneuve d’Asq* has been built close to Lille. Like *Cergy-Pontoise* this town has been conceived according to the general guidelines of these typically French projects, which are inspired by the concept of English new towns.

Lyon is an agglomeration of more than a million inhabitants with a long history. At the beginning of industrialisation, Lyon was one of the main centres of the textile industry, which declined later on. But after having converted the industrial areas, Lyon enjoyed an economic renewal. Due to the topography, the hinterland is highly contrasted: the mountainous western periphery is dominated by individual housing of high standard, whereas in the south and the east, social housing and industrial activities are more present.

The *Stuttgart* metropolitan area, more recently industrialized, was hardly influenced by the presence of car industry and electric goods industry. Many of the industries, big ones as well as a great number of middle-sized and small ones, are

⁵ In the frame of this project a software package called fractalyse was developed by Gilles Vuidel. It allows analyzing urban patterns by means of different methods of fractal analysis. Basic operations of image analysis like dilation have been integrated as well as different methods for estimating fractal parameters. If you want more information about fractalyse, please consult the website of the research team ThéMA: <http://thema.univ-fcomte.fr>, heading “Research teams” → “City, mobility, territory”. The software can now be downloaded from the website <http://fractalyse.org>

⁶ The investigations within the framework of the contract already mentioned were carried out by D. Badariotti (Strasbourg and Saarbrücken), I. Thomas and M.-L. De Keersmaeker (Brussels), C. Tannier and B. Reitel (Basel), G. Rabino and M. Cagliani (Milan). The paper refers mainly to the investigations carried out at ThéMA (Besançon) by L. Quiévreux and P. Frankhauser.

localized in smaller towns of the hinterland; some of them could be characterised as edge-cities. Since a couple of years efforts have been made to coordinate planning policy on the scale of the metropolitan area. Urbanisation has been canalized along valleys and transportation axes and by limiting the establishment of commercial areas and malls. Moreover land conservation plans been drawn up linking green areas on different scales has. Nevertheless the various towns remain in competition with each other and try to realize their own development policies, often influenced by economic actors.

The urban area of *Montbéliard* is a conurbation consisting of several towns of about 30,000 inhabitants. Their development was highly influenced by the presence of the Peugeot factory, which was established in the 1950's. The rapid development of the car industry necessitated the rapid construction of social dwellings. Most of them were constructed around the main factory and conceived, as most social housing in France at this period, according to the principles of the Le Corbusier School. Individual housing areas were often constructed in the vicinity of these zones. On the scale of the agglomeration, the urbanisation didn't follow a particular strategy but responded to a concrete urgent demand and to economic pressure. Nowadays local planning policy is trying to reassess the spatial structure of the area by specific measures and various master plans.

Finally, in the case of *Brussels* some outskirts in the Walloon Brabant will be taken into account. The main interest of this case lies in the well-known fact that sprawl was less controlled in Belgium than in other European countries (De Keersmaecker et al. 2003). In a similar way some parts of Bergamo were considered, which belongs to an Italian region where urban sprawl is rather important, due to the relative proximity to the Milan agglomeration on the one hand and the landscape amenities on the other.

4.2 The Results for the Built-Up Area

For all city centres of the big towns analysed, the patterns are rather uniform; the values of their *surface dimension* D_s lie in the range between 1.8 and 1.95. This is linked to the intensity of the land use, due to the high land prices. Districts situated in the immediate neighbourhood of the city centres often show similar values.

In urban areas where there are fairly big towns on the periphery of the main city, these towns have often rather low D_s -values. In the *Stuttgart* case, the D_s -values of the peripheral towns fall into a range of 1.61 to 1.73, except in the case of *Fellbach*, where urbanisation has been less controlled. For conurbations the situation is different since all the cities have more or less the same size and comparable functions; this holds e.g. for *Montbéliard* where all towns have a general value of ca. 1.8.

The values obtained for outskirts or particular types of districts on the periphery of great cities vary within a rather large range. The differences may easily be linked to the specific context in which they were constructed. Purely individual housing estates often show a D_s -values comparable to city centres. *This makes*

again evident that the information transcribed by fractal analysis has nothing to do with density.

Such patterns were observed in the east of *Lyon* (1.82 to 1.99!) but also in some districts in the west of *Lyon*. Comparable values exist for some peripheral districts and outskirts of *Stuttgart* but also for some outskirts of *Helsinki*. Such situations are often the result of regulations that impose a *constant rate of land-use all over a large zone*, and where no public or private service areas like commercial areas, schools etc. or leisure areas are established. Hence these zones are *rather uniform* in their spatial structure.

Less planned and constrained outskirts are often more contrasted and irregular. This is true for some outskirts in the south of *Brussels*, but also for *Lille* and for areas with less controlled progressive urbanisation in the *Stuttgart* region. Here we find rather low D_s -values from 1.50 to 1.74. Some results obtained for the *Bergamo* region confirm these observations.

Patterns that contain industrial areas are usually contrasted. This is true for the southern part of the *Lille* region, but also for areas in the south of *Lyon* or in the region of *Strasbourg*.

Contrasted patterns also occur when planners intend to integrate public space with different functions – and thus different sizes – in a town section. *Here the situation is completely different from that of homogenous individual housing estates*. Different types of architectural concepts pursue such objectives. Districts constructed according to the scheme of Le Corbusier have rather low D_s -values (1.54 to 1.77). Such districts were identified at *Montbéliard*, *Stuttgart*, *Lyon* and *Strasbourg*. Very comparable situations occur for the “new towns”. In the different districts of *Cergy-Pontoise* and *Villeneuve-d’Asq*, highly contrasted planned patterns were observed (1.60 to 1.73).

Some results, obtained for agglomerations near national borders, may complete these observations. In all cases no fundamental differences were observed even if there are nuances in some cases. In the *Strasbourg* region, residential outskirts show comparable values for both sides of the Rhine river. In the case of *Saarbrücken*, rather low D_s -values are found for German and French districts in general. Here one should take into account the fact that industrialisation has influenced urbanisation greatly, like in some parts of *Lille*. For the outskirts of *Basel*, the differences were more important: the French parts in particular are rather uniform, whereas the German and the Swiss estates have D_s -values of about 1.77 indicating a higher contrast.

4.3 The Results for the Urban Boundaries

As pointed out previously, the local planning policy often tends to act on town outlines. This section focuses on the morphology of these outlines. We refer to the correlation dimensions $D_b^{(tot)}$ of boundaries extracted after dilating by few steps the origin pattern (3 to 4 steps) which describes rather the fragmentation of the pattern as pointed out before. Sometimes we have also determined the dimension

of the boundary of the biggest cluster which should rather be interpreted as a measure for the tortuous character of the boundary.

The city centres are not of interest here and only the peripheral zones of agglomerations and outskirts will be considered. When comparing the periphery of agglomerations, considerable differences can be observed. The boundary dimension $D_b^{(tot)}$ for the western part of the Lille agglomeration for instance is 1.87. For the three persisting big clusters we have $D_b^{(clust)}$ -values of about 1.39. Hence the boundaries are rather tortuous and the ensemble of the clusters' boundary tends to cover intra-urban space in a rather uniform way. The comparison of these values with that of the *Stuttgart* region turns out to be interesting, since in the case of *Stuttgart* the planning policy tends to control sprawl strongly on the micro-scale of the town sections. Indeed the $D_b^{(tot)}$ -value for the main cluster of *Stuttgart* amounts to 1.80, and the $D_b^{(clust)}$ -value of the main aggregate is 1.27. The values are lower, the boundaries on the scale of the town sections are smoothed out, but the *Stuttgart* agglomeration also features other cases: as mentioned above, the local authorities in town of *Fellbach* did not impose such strong controls on urban sprawl: The $D_b^{(clust)}$ -value of the main cluster obtained after dilation is 1.45 and the $D_b^{(tot)}$ -value is 1.7. Hence the boundary is very tortuous and the ensemble of the intra-urban boundaries is more contrasted since $D_b^{(tot)}$ is lower.

On a more detailed scale it is interesting to consider the outline of districts situated on the periphery of towns as well as that of outskirts. For the same reasons as discussed before, some examples of the *Stuttgart* region shall be presented. Settlements dominated by individual housing areas show rather low boundary dimensions for the main cluster (about $D_b^{(clust)} = 1.26$), but high values for $D_b^{(tot)}$ (1.80). This can be explained by the fact that these outskirts were constructed according to a logic of housing estates where planning defined a rectilinear border and imposed rather rigid constraints about plot size. Thus intra-urban non-built-up space is distributed rather homogeneously.

For the *Lyon* and the *Lille* agglomerations, the results are rather different. On the eastern and southern periphery of *Lille*, the $D_b^{(clust)}$ -values lie within a broad range (1.33 to 1.55) and the $D_b^{(tot)}$ -values indicate a rather homogeneous pattern, but other town sections show a higher inner contrast. All these town sections seem to be the result of a progressive, less controlled growth process and some of them show a high mix of residential and industrial areas including public buildings like schools. Thus the patterns remind us more of a teragon ($D_b = 1.50$), in particular when their intra-urban pattern tends to a uniform distribution of build-up space. In this context two examples analysed in the *Bergamo* region of recently developed mainly residential towns are of interest, too: they show highly dendritic boundaries (1.43 to 1.46). Obviously no restrictions have been imposed for smoothening the urban boundary.

These results show again that earlier built town sections, particularly if they are the result of a progressive and less controlled urbanisation process, where industrial activities played an important role have more irregular boundaries. A comparable situation arises for recent residential zones resulting from a more spontaneous growth process. The differences obtained for the peripheral towns in the Stuttgart region confirm this result. On the other hand, the stronger planning

policy in the Stuttgart region tends in general to limit sprawling on the scale of the districts.

Finally, the question of a possible influence of the national context on the morphology of outlines will be considered again. The comparison of the outlines of different parts of the *Saarbrücken* region, which was carried out for a couple of French and German districts, did not yield particular differences. In the case of *Basel*, the main clusters of the Swiss outskirts have $D_b^{(clust)}$ -values resembling to the values obtained for the Stuttgart region (1.27), the German *Lörrach* has a higher value (1.37). For the French part the tortuousness is higher (1.47), this could correspond to a slightly looser control.

Finally, the particular case of strongly planned districts like those of *Cergy-Pontoise* is discussed. The values observed fall into the same range as those observed for the outskirts of Stuttgart (about 1.29), matching a tendency to smoothen boundaries. The total boundary however, has a rather low value ($D_b^{(tot)} = 1.7$) that corresponds to the very hierarchical spatial organisation of the intra-urban space. This situation resembles the morphology of the Sierpinski carpet of figure 6a, which is very contrasted inside but has a smooth boundary. It may be surprising that in contrast to Cergy-Pontoise, the new town of *Villeneuve d'Asq* on the periphery of Lille has a rather dendritic, tattered outline ($D_b^{(clust)} = 1.38$). But this new town generally pursues a less strict planning concept than Cergy.

5 Urban Pattern Morphogenesis as Self-Organisation Process

Fractal analysis of a great number of urban patterns made evident that the fractal power law is generally rather well suited for describing their morphology. This does not exclude that the scaling exponents may vary when passing from characteristic distance ranges to other ones, e.g. from the scale of town districts to that of the agglomeration. Fundamental deviations are mainly observed for less urbanized areas, for very irregular patterns or when passing from the scale of town sections to that of whole the agglomeration. Moreover the scale of buildings and courtyards shows typically non-fractal behaviour.

We may conclude that urbanisation generates obviously a particular type of pattern which follows an *internal order principle*. This order may be described by morphological parameters, the different *scaling exponents* (fractal dimensions). Since this type of pattern organisation is not the product of a strict planning policy but rather the result of the interaction of a great number of decision makers, we may interpret urban pattern morphogenesis as *self-organisation process*, in the sense of *synergetics* (Haken 1978)⁷.

⁷ The question whether urban pattern formation may be considered as self-organisation process has been considered by several authors. Let us remind the early attempts of Allen and Sanglier 1981; Pumain et. al. 1989; Haag and Dendrinis 1983, or Weidlich and Munz 1990. The numerous models based on cellular automata for simulating urban

On the other hand we observed that the values of the morphological parameters may depend on certain conditions: local planning policy or particular architectural concepts influence the parameter values; this holds for the micro-scale of districts as well as for the scale of metropolitan areas. From a synergetic point of view these conditions may be interpreted as “control parameters”. Other parameters, like national context, turned out to be less important: more or less regular patterns exist in all countries. Topographic constraints like mountains may influence urban morphogenesis on the scale of the agglomeration, but eventually urban growth along transportation axes may generate the same type of patterns as if an agglomeration is situated at the intersection of several valleys.

Radial analyses of urban areas showed that local deviations from a pure fractal law are considerably more important in the curves of scaling behaviour for low urbanized areas; fractal order seems to be less present. Recent investigations of the Walloon region (Thomas and Frankhauser 2005), based on correlation analysis, confirmed these observations.

Hence we defined like in physics an “order parameter” which allows evaluating to what extent an urban pattern corresponds to a fractal organisation (Frankhauser 2000). We computed this parameter for several agglomerations for which we had cartographic data for different time periods at our disposal. This allows observing whether the spatial organisation of the patterns corresponds more to a fractal order in course of urbanisation. The used data sets are comparable for each time series.

The definition of the order parameter referred directly to the previously introduced curves of scaling behaviour. As pointed out these curves allow distinguishing local deviations from fractal laws from real changes of fractal behaviour. In order to separate strictly both the phenomena, we resorted to Gaussian smoothing of the curves of scaling behaviour. By choosing appropriate values for smoothing it is possible to eliminate the local fluctuations, but to conserve the long range changes due to real changes of fractal behaviour. For a discrete data set the formula for smoothing reads

$$\tilde{\alpha}_i(\varepsilon_i) = \sum_{j=1}^n \alpha(\varepsilon_j) \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(\varepsilon_j - \varepsilon_i)^2}{2\sigma^2}}$$

The smoothed value $\tilde{\alpha}_i(\varepsilon_i)$ is the mean value of all weighted values $\alpha(\varepsilon_j)$ lying within a range defined by the σ -value. Besides the purely technical aspect we may give a statistical interpretation to this procedure. If we assume that we have a virtual set of similar towns which show the same morphological characteristics, we may suppose that for each distance ε_i the $\alpha(\varepsilon_j)$ -values fluctuate around a value, $\tilde{\alpha}_i(\varepsilon_i)$ peculiar for this distance. We assume now that deviations from this value are randomly distributed according to a Gaussian law and we suppose that the occurring values are just identical to the $\alpha(\varepsilon_j)$ -values observed in our patterns in the

growth (early examples cf. e.g. Batty 1981; White and Engelen 1993) as well as the use of multi-agent models correspond to an equivalent reasoning (Schweitzer 2003).

vicinity of distance ε_i . Then $\tilde{\alpha}_i(\varepsilon_i)$ becomes just the typical mean value for distance ε_i .

Following the usual definition, we introduce now two different types of variances. According to our data model we may define for each distance ε_i local variances of our virtual data around the typical mean value:

$$\sigma(\varepsilon_i) = \frac{1}{n} \sum_{j=1}^n \left(\alpha(\varepsilon_j) - \tilde{\alpha}(\varepsilon_i) \right)^n$$

where n is the number of data in the data range defined by the Gauss convolution. Moreover we introduce now the mean variance of these local variances for whole the range of distances $i = 1 \dots m$:

$$\sigma_{res} = \frac{1}{n \cdot m} \sum_{i=1}^m \sum_{j=1}^n \left(\alpha(\varepsilon_j) - \tilde{\alpha}(\varepsilon_i) \right)^2$$

On the other hand we consider now the variance with respect to a *global mean value* of all $\alpha(\varepsilon_i)$ -values. We may interpret this variance as a measure for the deviations of empirical data with respect to a *constant fractal dimension* all over the range of distances ε_i considered:

$$\sigma_{tot} = \frac{1}{n \cdot m} \sum_{j=1}^n \sum_{i=1}^m \left(\alpha(\varepsilon_j) - \bar{\alpha} \right)^2 = \frac{1}{n \cdot m} \sum_{j=1}^n m \left(\alpha(\varepsilon_j) - \bar{\alpha} \right)^2 = \frac{1}{n} \sum_{j=1}^n \left(\alpha(\varepsilon_j) - \bar{\alpha} \right)^2$$

Both these variance measures allow us now to introduce a parameter which corresponds by its definition to the so-called *correlation ratio* in statistics. This parameter allows measuring the intensity of a relation between two variables when having different data sets at our disposal. Then $\sigma(\varepsilon_i)$ are the variances in each data set, whereas σ_{res} is called *residual variance* or *intra group variance* and describes the mean variance in the ensemble of all the data sets. On the other hand σ_{tot} is the total variance which respect to the mean value, when putting together all data sets. The more the relation is strong the more the variances $\sigma(\varepsilon_i)$ will be small, since to each value ε_i more or less one unique $\alpha(\varepsilon_i)$ -value would correspond. On the contrary if the variances $\sigma(\varepsilon_i)$ are important and their mean σ_{res} approach the total variance the relation is weak. Hence the correlation ratio is defined in the following way:

$$\eta^2 \equiv 1 - \frac{\sigma_{res}}{\sigma_{tot}}$$

According to the previous discussion, η^2 tends to *zero*, if the relation is weak, i.e. if $\sigma_{res} \rightarrow \sigma_{tot}$. On the contrary for a strong relation η^2 tends to *one*, since $\sigma_{res} \rightarrow 0$. In our case the situation is similar: if important variations of the $\alpha(\varepsilon_i)$ -values occur, they will dominate the feature of the curve of scaling behaviour and $\sigma_{res} \rightarrow \sigma_{tot}$. Then no particular fractal behaviour is observed. On the contrary, when the $\alpha(\varepsilon_i)$ -values do not deviate much from the smoothed curve, $\sigma_{res} \rightarrow 0$. This condition is particularly fulfilled, if all the $\alpha(\varepsilon_i)$ -values are equal and thus identical to their mean value. This corresponds to a perfect fractal structure.

Hence the parameter fulfils the requirements for an order parameter as defined in physics: the value tends to zero if the considered structure is absent, and tends to one when the structure is perfect.

Figure 14 shows the time evolution of the fractal order parameter for some agglomerations⁸. We get aware that for the large metropolitan areas like *Berlin* or *Munich*, the fractal order increases at least on the long run. This holds too for the small town of *Lons-le-Saunier*, situated in the East of France, where a slight decrease is observed for the most recent period. For the *Montbéliard* agglomeration we observe an interesting phenomenon: the η^2 -value declines for the historical centre in a first period and increases not much during the second period, whereas for the *Peugeot factory* an important increase is observed. *Audincourt* belonging to the same conurbation remains rather on a low level, too. This makes evident that urbanisation is dominated by the dynamics of the Peugeot enterprise: the new urban areas are constructed in the vicinity of this factory.

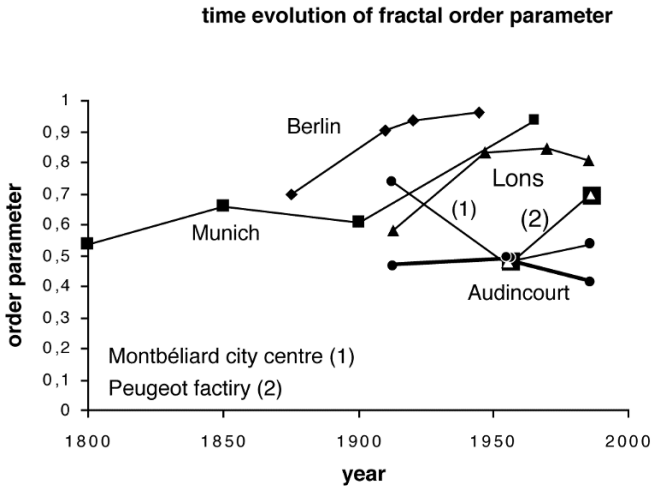


Fig. 14. Evolution of the fractal order parameter for some agglomerations (cf. text).

These results seem to be in concordance with the observations of F. Schweitzer and J. Steinbrück (1998, 2002), who found that the hierarchy of the clusters in the metropolitan zones of Munich and Berlin, i.e. another fractal characteristic of urbanized areas, becomes more regular in course of urbanisation.

⁸ The data sets used for Munich and Berlin are issued from aggregated cartographic representations conceived by the Städtebauliches Institut (University of Stuttgart), whereas the other data bases have been developed at Théma on the base of topographic maps (1:50,000).

5 Conceptual Conclusions

The fact that urbanization tends to generate fractal structures should of course incite reflecting about the underlying socio-economic processes. A detailed analysis of the complex interactions between the different kinds of actors contributing to these dynamics is of course not possible in the present contribution. However we will come back to some fundamental properties of fractals which may qualitatively be linked to some well-known topics of social demand. It is known that one important argument for choosing outskirts as residence is – besides the lower lot prices – the quality of life which is associated to these sites. Such sites allow living in an individual house with garden, in a calm and reassuring environment, embedded in a nice landscape. On the other hand the accessibility of these ancient rural regions from the central town has been progressively improved what allows to profit also from the services offered by city centres like up-market shopping, cultural activities etc. Since the road network covers space more and more uniformly, developer may construct housing estates nearly everywhere. Hence these zones risk becoming victims of their success: e.g. the fractal dimension of the metropolitan area of Strasbourg increased in the short time period from 1986 to 1996 from $D_s = 1.80$ to $D_s = 1.86$ and hence tend to uniformity. This process makes on the long run fragile the amenities residents searched and which play moreover an important role for whole the urbanised areas like large natural reserves.

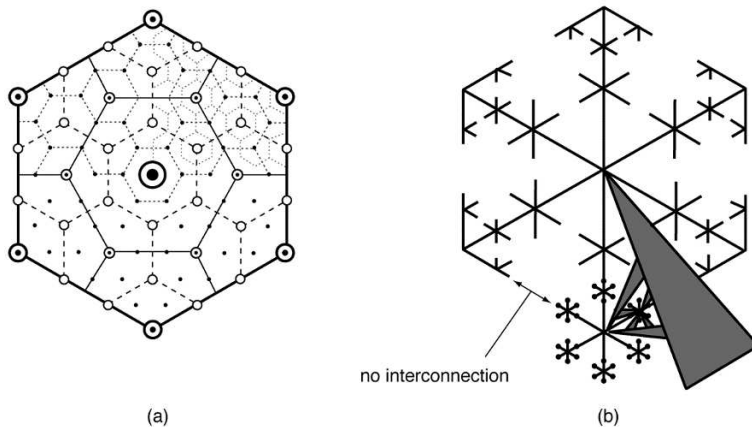


Fig. 15. The spatial system of central place theory according to Christaller (a) and a multifractal hexagonal network (b). The outer parts have been cut off in order to make evident the comparison with the Christaller scheme, excepted for the bottom part. By renouncing to the uniformity principle the fractal settlement system allows to concentrate population near transportation axes. A connected network of green areas may be preserved as illustrated at the bottom. The scaling logic of the pattern makes possible creating leisure areas near to urbanized areas

Indeed uniform distributions can *never offer a diversified range of amenities*. If homogeneously built-up residential areas offer an individual garden for each house, they provide no *public space* for events, which necessitate free areas on a different scale. Homogeneous sprawl patterns like in the urban plan of Los Angeles potentially generate traffic flows since the local range on offer is not diversified enough. But not only sprawling patterns are susceptible to generating traffic flows, this holds for compact cities too, since residents will try to reach open space for leisure activities. In fact it is – even from a purely climatologic point of view – no longer possible to imagine huge population concentrations like those of important metropolis confined in a densely populated area.

Based on the consuming behaviour of urban amenities Christaller (1933) proposed a hierarchical system of central places where high level services are concentrated in few large towns. Medium sized towns offer lower ranked services, whereas small towns are only equipped with basic services. However the spatial pattern proposed remains homogeneous (Fig. 15a). Figure 15b shows a fractal system, based, too, on such a hierarchy but which cuts off, according to a multi-scale logic, empty space. Such a spatial organisation offers the chance of having on the one hand zones of high concentration near urban amenities (shopping, services), e.g. in the vicinity of public transportation networks, and on the other hand more diluted zones. So it becomes possible to maintain a social mix by means of a higher local variety of densely and less densely populated zones and, on the other hand, to preserve huge empty zones in the neighbourhood of urbanized areas, which may be imagined as natural reserves, agricultural zones or simply leisure areas offering “rural” amenities.

There exist various planning concepts and real world situations which are close to such logic, like the concept of “*urban villages*” which tries introducing operational sub-centres with pertinent service amenities in order to reduce traffic flows to the main centre (Fouchier 1995). The low D_s -values obtained for Cergy-Pontoise correspond to the objective of diversifying the range of public space on offer. Such low values were also observed for districts constructed according to concepts from the School of Le Corbusier with comparable planning intentions⁹ (cf. the plan of Hilberseimer Fig. 16a). A similar situation occurs in the case of the green area concept of Stuttgart, but in fact also in the real world pattern of Copenhagen (Fig. 16 c), since the fingers are not homogeneously filled up with built-up space.

⁹ It should be emphasized that the social problems occurring in these town sections are related to the effect of “crowding”, due to the heights of buildings as well to their uniform architectural realization. These effects cannot of course be analysed in our context where we restricted the analysis to the occupation of the surface, without taking into account the height of buildings. However three-dimensional analysis is possible when one includes data about the height of buildings or the population distribution on a detailed scale (Frankhauser 1998).

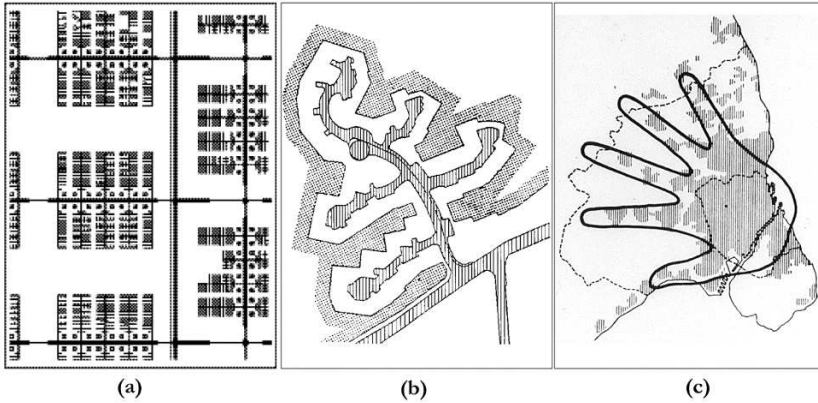


Fig. 16. a) the plan of Hilberseimer for the reconstruction of Chicago follows a strong fractal logic. b) the architect Schöfl proposed the interpenetration of green areas with housing in the fringe of towns. c) the “finger plan” of Copenhagen for avoiding uncontrolled sprawl: leisure areas are close the urbanized areas which are localized on public transportation network axes

Multi-scale concepts may also be of interest on the micro-scale of town sections. Instead of rounding up the boundaries, fractal-like outlines of residential areas may be imagined: the consequent lengthening of the outline lets more house-owners benefit of the situation on the edge of the settlement which seems to be rather appreciated. The model proposed by the architect Schöfl (Fig. 16b) illustrates how such arrangements could be developed. *This could help to avoid sprawl on a larger scale, which tends to generate homogeneous land occupation and thus weakens the quality of the landscape. Accepting a “local controlled sprawl” could be beneficial in terms of avoiding sprawl on a larger scale and encourage sustainability.*

When interpreting the results it becomes obvious that a fractal approach to urban patterns helps us to make evident that urban patterns follow a particular type of spatial organisation, despite of their irregular aspect. Besides the purely descriptive approach, fractal modelling allows illustrating important features of these patterns and seems to provide a new approach for managing the consequences of the new lifestyle, which tends to claim good access to different kinds of both urban and rural amenities, and at the same time helps to reduce the risks of a diffuse sprawl which tends to weaken the environmental quality and to generate more and more traffic flows.

Developing a well-founded methodology for rendering operational this approach should be an important goal for future research.

Acknowledgments

We are very indebted to the French Ministry of Planning, Transport, Housing, Tourism and Sea for having supported the previously presented research activities as part of the program “ville émergente”. We want to thank all participants of the above project as well as all the local authorities and the Städtebauliches Institut that provided the used data. Moreover we are grateful to Gilles Vuidel for having developed the pertinent computer program “fractalysé”.

References

- Allen P, Sanglier M (1981) Urban evolution, self-organisation and decision making. *Environment and Planning*, **13**:167-183
- Batty M (1991) Generating urban forms from diffusive growth. *Environment and Planning A* **23**: 511-544
- Batty M, Longley PA (1986) The fractal simulation of Urban Structure. *Environment and Planning A* **18**: 1143-1179
- Batty M, Longley P (1994) *Fractal Cities. A Geometry of Form and Function*. Academic Press, San Diego, p. 394
- Batty M, Kim SK (1992) Form Follows Function: Reformulating Urban Population Density Functions. *Urban Studies* **29**: 7: 1043-1070
- Batty M, Xie Y (1996) Preliminary evidence for a theory of the fractal city. *Environment and Planning A* **28**: 1745-1762
- Benguigui L, Czamanski D, Marinov M, Portugali J (2000) When and where is a city fractal? *Environment and Planning B* **27**:4: 507-519
- Cavallières J, Frankhauser P, Peeters D, Thomas I (2004) Where Alonso meets Sierpinski: an urban economic model of a fractal metropolitan area. *Environment and Planning A* **36**: 1471-1498
- Christaller W (1933) *Die zentralen Orte in Süddeutschland*. Fischer, Jena
- De Keersmaecker, ML, Frankhauser P, Thomas I (2003) Using Fractal Dimensions for Characterizing Intra-Urban Diversity: The Example of Brussels. *Geographical Analysis* **35**: 310-328
- Dubois-Taine G, Chalas Y (eds) (1997) *La ville émergente*. Edition de l’Aube, Paris
- Dubois-Taine G (coord. by) (2002) *La ville émergente, Résultats de Recherche*. Report PUCA, Ministère de l’Equipement, des Transports, du Logement, du Tourisme et de la Mer, Paris, p 205
- Dubois-Taine G (ed) (2004) *European Cities – Insights on outskirts*, Report COST Action 10 Urban Civil Engineering. Vol 1: From Helsinki to Nicosia. Eleven Case Studies & Synthesis, Brussels
- Fouchier V (1995) La densification: une comparaison internationale entre politiques contrastées. *Les Annales de la Recherche Urbaine* **67**: 94-108
- Frank H (1987) Idee, Prozess, Ergebnis. Die Reparatur und Rekonstruktion der Stadt. In: *Die Überwindung der Stadtbaukunst, Internationale Bauausstellung Berlin*. Fröhlich und Kaufmann, Berlin

- Frankhauser P (1988) Fractal aspects of urban systems. Beiträge zum ersten Internationalen Symposium des SFB 230 Teil 1. SFB 230, Stuttgart, pp 67-76
- Frankhauser P (1994) La fractalité des structures urbaines. Collection Villes, Anthropos, Paris
- Frankhauser P (1998) The fractal approach. A new tool for the spatial analysis of urban agglomerations. Population: an English selection, special issue New Methodological Approaches in Social Sciences, pp 205-240
- Frankhauser P, Genre-Grandpierre C (1998) La géométrie fractale, un nouvel outil pour évaluer le rôle de la morphologie des réseaux de transport public dans l'organisation spatiale des agglomérations. Les Cahiers Scientifiques du Transport **33**: 41-78
- Frankhauser P (2000). La fragmentation des espaces urbains et périurbains - une approche fractale. Structures des villes, entreprise et marchés urbains, l'Harmattan, collection Emploi, Industrie et Territoire, Paris
- Frankhauser P (coord. by) (2003) Morphologie des villes émergentes en Europe à travers les analyses fractales. Research Report, Université de Franche-Comté, UMR 6049 THEMA, p 240
- Frankhauser P (2004) Comparing the morphology of urban patterns in Europe – a fractal approach. European Cities – Insights on outskirts, Report COST Action 10 Urban Civil Engineering, vol 2: Structures, Borsdorf and Zembri, Brussels, pp 79-105
- Frankhauser P, Tannier C (2005) A multi-scale morphological approach for delimiting urban areas. Working-paper presented at the Conference “Computers in Urban Planning and Urban Management”, London
- Gouyet JF (1992) Physique et structures fractales. Masson, Paris
- Haag G, Dendrinos DS (1983) Towards a stochastic dynamical theory of location: a non-linear migration process. Geographical Analysis **15**: 269-286
- Haken H (1978) Synergetics: An Introduction. Springer-Verlag, Berlin
- Humpert K, Brenner K, Bohm H (1991) Großstädtische Agglomerationen,- ein globales Problem. Natürliche Konstruktionen, vol 1, SFB 230, Stuttgart
- Humpert K, Brenner K (1992) Das Phänomen der Stadt als fraktale Struktur. In: Das Phänomen der Stadt. Berichte aus Forschung and Lehre, Städtebauliches Institut, Universität Stuttgart, pp 223-269
- Humpert K, Brenner K, Becker S (2002) Fundamental Principles of Urban Growth. An Analysis of 57 cities. Müller und Busmann, Wuppertal
- Le Bras H (1993) La planète au village. Editions de l'Aube, Paris
- Longley P, Mesev V (2000) On the Measurement and Generalisation of Urban Form. Environment and Planning A **32**: 3: 471-488
- Mandelbrot B (1983) The Fractal Geometry of Nature. Freeman, San Francisco
- Pumain D, Saint Julien T, Sanders L (1989) Ville et auto-organisation. Economica, Paris
- Rémy J (1994) La ville: réseau alvéolaire et mobilité spatiale. In: Pellegrino P (ed) Figures architecturales, formes urbaines. Anthropos, Paris
- Schöfl G (1986) Minimalnetze. arcus **2**
- Schweitzer F, Steinbrick J (1998) Estimation of Megacity Growth. Simple rules versus complex phenomena. Applied Geography **18**:1: 69-81
- Schweitzer F, Steinbrick J (2002) Analysis and computer simulations of urban cluster distribution. In: Humpert K, Becker S, Brenner K (eds) Fundamental Principles of Urban Growth. Müller und Busmann, Wuppertal, pp 142-157
- Schweitzer F (2003) Brownian agents and Active Particles. Springer, Berlin

- Shen G (2002) Fractal dimension and fractal growth of urbanized areas. *International Journal of Geographical Information Science* **16**:5: 519-437
- Sievert T (1997) Zwischenstadt. *Bauwelt Fundamente* **118**: 193
- Tannier C, Pumain D (2005) Fractals in urban geography: a general outline and an empirical example. *Cybergeo* **307**: 22 <http://www.cybergeo.presse.fr>
- Thomas I, Frankhauser P (2005) Fractal dimension, morphology of the built-up surfaces and land-use planning. Paper presented at the 45th Congress of the European Regional Science Association, Amsterdam
- Weidlich W, Munz M (1990) Settlement formation. Part 1: A Dynamic Theory; Part 2: Numerical Simulation. *The Annals of Regional Science* **24**: 83-106, 177-196
- White R, Engelen G (1993) Cellular Automata and Fractal Urban Form: A cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A* **25**: 1175-199
- White R, Engelen G (1994) Urban Systems Dynamics and Cellular Automata: Fractal Structures between Order and Chaos. *Chaos, Solitons and Fractals* **4**: 4: 563-583

Cartographic Data Sources for Figures

<i>Besançon</i>	GIS-data (Agence d'urbanisme de l'agglomération bisontine).
<i>Brussels</i>	GIS-data (Centre Informatique pour la Région Bruxelloise).
<i>Lille</i>	GIS-data (Communauté urbaine de Lille).
<i>Montbéliard</i>	GIS-data (Agence d'urbanisme du CAPM) and conception by ThéMa, based topographic maps.
<i>Stuttgart</i>	detailed: conception by ThéMa, based topographic maps, coarse-grained: conception by Städtebauliches Intitut (University of Stuttgart).

The Dynamics of Complex Urban Systems: Theory and Application of the STASA-Model within the Scatter Project

Günter Haag, Jan Binder

Steinbeis-Transfer Centre Applied Systems Analysis (STASA), Stuttgart, Germany

Abstract

In order to understand current settlement pattern and its future development the essential characteristics of the interrelated dynamics of the transport system, the urban/regional settlement pattern, demographic effects and the impacts of different policy measures and external effects have to be considered in an interdisciplinary approach. The modelling of the spatio-temporal patterns of a system consisting of different sub-models (population, transport, production, etc.) is used to demonstrate the necessity to link different theoretical frameworks. External shocks may sometimes require a modification of the system under consideration in the sense that new dynamic variables appear or previously useful variables disappear.

Policy advises on the base of different scenarios are useful in a planning context in the attempt to shape the future taking into account different scenario results, instead of just adapting to what may emerge. This procedure is especially important when essential parts of the system cannot be controlled by policy measures and depend on exogenous factors.

For the region of Stuttgart, the model framework was used within the SCATTER project. Four different policy scenarios: reference scenario (no additional policy measure applied); fiscal measures; regulatory measures applied to companies and increase of car use costs will be simulated in an recursive procedure. The results of the different policy measures will be discussed and critically evaluated with respect to urban sprawl and sustainable development aspects. The developed framework can also be used as simulation tool for the estimation of secondary induced traffic impacts.

1 Introduction

Urban sprawl is widely spread over Europe. It induces high level of car use and, usually, congestion on roads giving access to city centres. To limit the damages caused by urban sprawl in terms of congestion, air pollution and energy consumption, numerous European cities are implementing suburban public transport services, such as heavy or light rail. But by improving the accessibility, they create an incentive for a new wave of urban sprawl. Therefore, in parallel with these

new public transport services, accompanying measures have to be elaborated and implemented, in order to prevent, mitigate or control urban sprawl. The proposed project tackles this issue in which land use and transport are closely mixed up. The key aim of the project is to promote sustainable development. In addressing transport, land-use and environment in urban context, SCATTER covers the most important threats to the well-being of the majority of European citizens (SCATTER 2001).

The results presented in this paper are a small part of the SCATTER project funded by the European Commission under the « Energy, Environment and Sustainable Development Programme » of the 5th Framework Programme, Key Action 4 : City of Tomorrow and Cultural Heritage, Contract number: EVK4-CT-2001-00063 (SCATTER 2004). Details of the SCATTER project are discussed by Batten et al. (2003) and Gayda et al. (2003a, 2003b).

In this paper the modelling framework as well as different simulations of policies aiming to reduce or control sprawl in the case city Stuttgart are presented. This requires an integrated modelling of the different subsystems (transport-, urban and regional-subsystem) and its interactions, without using any equilibrium assumption. Especially the redistribution of traffic flows caused on migration decisions and the effects of commuting related to different policy measures will be investigated.

2 The Stuttgart Case City

The Stuttgart Region and the case city of Stuttgart is situated in the south-west of Germany and covers 6 districts (NUTS 3) called Boeblingen, Esslingen, Goeppingen, Ludwigsburg and Rems-Murr, and the City of Stuttgart with a total of 179 (Gemeinden) communities (Fig. 1). The overall population of the Stuttgart Region is about 2.6 million inhabitants. Together with its state capital, the City of Stuttgart, it represents the economic and cultural centre of the state of Baden-Wuerttemberg. With an area of 3.700 km² this region is one of the most densely populated regions of Germany (> 700 inhabitants/km²).

The urban development is almost uniformly spread over the whole area. This is reflected in the location of many medium size and big municipalities (sub-centres) organised almost uniformly around the City of Stuttgart (see Fig. 2).

There are about 589.000 inhabitants concentrated in the City of Stuttgart (urban centre). This corresponds to 22% of the total population of the Stuttgart Region. Taking into account the adjacent communities (Sindelfingen, Boeblingen, Esslingen, Leonberg, Leinfelden-Echterdingen, Ludwigsburg, Schorndorf) within a small circle (15 Km) around the centre of the City of Stuttgart about 38% of the total population can be found (urban zones). Both the City of Stuttgart and its neighbouring communities are densely populated.

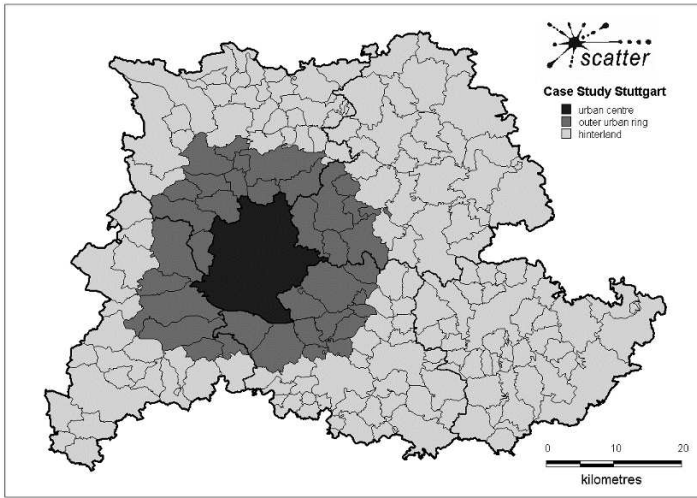


Fig. 1. Urban definition of case study Stuttgart

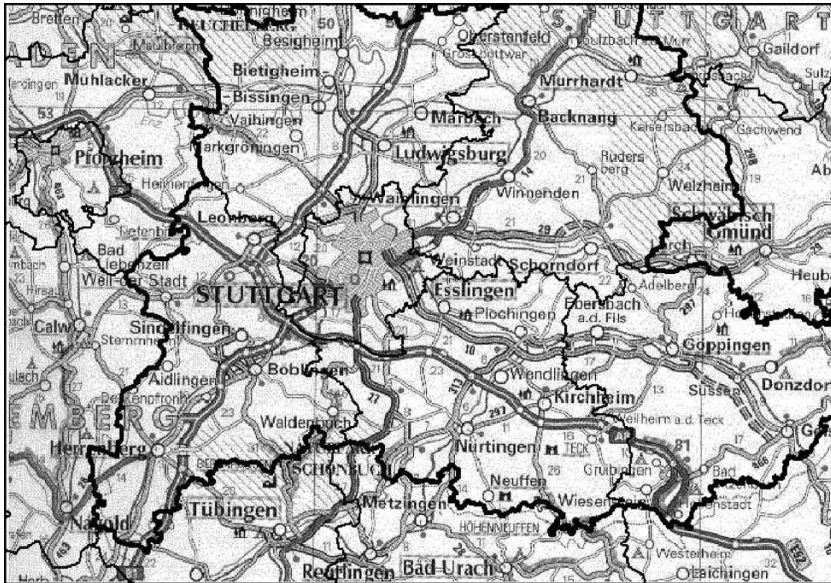


Fig. 2. Map of the Stuttgart Region (120 km x 80 km)

3 The STASA Land-Use and Transport Modelling Framework

The commuter flows (VIGONI 2001) and the migration flows are modelled via the master equation framework. In order to analyse both inter- and intra-regional flows the STASA-transport/land-use model (Weidlich and Haag 1988, 1999; BMBVW 1999; Haag and Grüzmann 2001; Haag 2002; Binder et al. 2003) has to be modified. In the following, a rather short description of the general modelling framework is presented.

Investments into the transport sector and communication networks improve the accessibility and attractiveness of suburban areas. This may lead to a redistribution of migration flows and traffic flows and is discussed as one possible reason for urban sprawl. The quantitative treatment of those nested processes of the different subsystems (transport-, population-, communication-subsystem) and its interactions require an integrated modelling.

On the one hand the dynamics on the macrolevel - i.e. the development of the traffic subsystem and of the urban/regional subsystem - is determined by the behaviour of the individuals on the micro-level, on the other hand "attractivity" differences between the spatial units (traffic cells), which depend on the macro-variables, influence the decisions of the individuals as well. Apart from rational motives of the actors several elements of uncertainty, e.g. irrational behaviour as a result of insufficient information, have to be taken into account. Hence, the description of decision processes is based on a stochastic and dynamical decision model within the master equation approach.

The traffic subsystem as well as the urban/regional subsystem form a complex intertwined system. Its dynamics take place on different time scales but are modelled making use of the same principles:

1. The daily traffic flows in the region of Stuttgart are the result of very quick decision processes of the actors to realize a trip between two traffic cells (origin - destination) with a special purpose. Decision processes for a certain destination, the moment of the setting out, the mode of transportation, the choice of the route etc. take place on a very short time scale.
2. The development of the urban/regional subsystem (e.g. spatial population distribution) is a process on a long-term time scale. The population distribution changes because of migration acts of individuals. The equations of motion which describe the migratory behaviour contain transition rates, i.e. migration flows between the cells. These flows depend on accessibility measures (coupling to the transport subsystem) and "attractivity" differences as a result of different regional advantages.

The total number of in- and out-commuters of communities has increased steadily in Germany for the last twenty years. In particular, this trend appears for far distance commuting. Nevertheless the willingness to commute diminishes sharply at a time distances larger than approximately 45 minutes (Johansson et al. 2001). Commuting offers the possibility to choose a new home by retaining the

workplace and vice versa. This is often combined with the acquisition of real estate or the improvement of working conditions by retention of the place of work without simultaneously losing the advantages of an existing residence (Steierwald and Kühne 1993).

The organisation of the traffic network is of crucial importance for commuting, and thus for urban sprawl, because the decision to commute depends among others on accessibility measures. Commuter flows represent an essential part of traffic flows in the rush hours. Considering this, it is obvious that changes in the socio-economic environment and investments in the traffic infrastructure has an impact on the number of commuter and distance of commuter flows, with feedback effects to the housing market.

The imbalance between labour demand and labour supply on the level of communities is partially compensated by the commuter dynamics. The development of regional employment, the regional gross wage payment and parts of the community revenues are strongly dependent on the spatial distribution of commuter flows (Binder et al. 2001).

Changes in the location of population and workplaces can be classified as a slow adjustment process. This means that commuting patterns should be expected to adjust itself to new conditions on a much faster time scale than household location pattern. Moreover econometric studies indicate that the location of firms adjusts faster than the location of households, and that workplaces tend to follow households (Mills and Carlino 1989, Holmberg et al. 2001, Johansson 2001).

The “empirical data base” consists of traffic flows (all modes) between the traffic cells, as well as population numbers on a yearly base between all communities and data on the traffic network.

These data base is used to estimate the system parameters of the transport and urban/regional subsystems, e.g. “transport attractivities” and “distance” (or resistance) parameters for the transport subsystem. In a further step these “attractivities” have been connected with appropriate macro-variables (key-variables) making use of a multiple regression.

The calibrated integrated model will be applied to the region of Stuttgart (Haag and Binder 2001). Different scenarios will be simulated and discussed with respect to the purpose of urban sprawl.

3.1 The Stuttgart Land Use and Transport Model Design

The population distribution is denoted by $\vec{n} = \{n_1, \dots, n_i, \dots, n_L\}$, where n_i is the number of individuals (households, persons) living in community i . n_i will be modified by the decisions of the people to commute between community i and any one of the other communities. Therefore, the population distribution \vec{n} is connected via commuter related activities or migration events with individual decision processes.

Let $P(\vec{n}, t)$ be the configuration probability to find a certain population distribution \vec{n} at time t , taking into account the complicated interactions. Of course this probability $P(\vec{n}, t)$ must satisfy the normalisation condition

$$\sum_{\vec{n}} P(\vec{n}, t) = 1 \quad (1)$$

where the sum extends over all possible population configurations \vec{n} .

The temporal evolution of the probability distribution $P(\vec{n}, t)$ can be described by the master equation (Weidlich and Haag 1984; Haag and Weidlich 1984; Haag 1989)

$$\frac{d}{dt} P(\vec{n}, t) = \sum_{\vec{k}} F_t(\vec{n}, \vec{n} + \vec{k}) P(\vec{n} + \vec{k}, t) - \sum_{\vec{k}} F_t(\vec{n} + \vec{k}, \vec{n}) P(\vec{n}, t) \quad (2)$$

where the sum on the right hand side extends over all \vec{k} with non vanishing configurational transition rates $F_t(\vec{n} + \vec{k}, \vec{n})$ and $F_t(\vec{n}, \vec{n} + \vec{k})$. Hereby the transition rate $F_t(\vec{n} + \vec{k}, \vec{n})$ (transition probability per unit of time) specifies the transition from any population distribution \vec{n} to a neighbouring distribution $\vec{n} + \vec{k}$.

The master Eq. 6 has a very direct and intuitively appealing interpretation. The change in time of the configuration probability $dP(\vec{n}, t)/dt$ is due to two effects of opposite direction: first to the probability flux from all neighbouring configurations $\vec{n} + \vec{k}$ into the considered configuration \vec{n} namely $\sum_{\vec{k}} F_t(\vec{n}, \vec{n} + \vec{k}) P(\vec{n} + \vec{k}, t)$ and second to the probability flux out of the configuration \vec{n} into all neighbouring configurations $\vec{n} + \vec{k}$, namely $\sum_{\vec{k}} F_t(\vec{n} + \vec{k}, \vec{n}) P(\vec{n}, t)$. Consequently, the master equation represents a balance equation for probability fluxes. The transition rates in the master equation are directly associated with the evolution of the conditional probability.

The transition rate $F_t(\vec{n} + \vec{k}, \vec{n})$ from the population distribution \vec{n} to the neighbouring distribution $\vec{n} + \vec{k}$ is the sum of all the contributions $F_{ij}^\alpha(\vec{n} + \vec{k}, \vec{n})$:

$$F_t(\vec{n} + \vec{k}, \vec{n}) = \sum_{\alpha=1}^A \sum_{i,j=1}^L F_{ij}^\alpha(\vec{n} + \vec{k}, \vec{n}) \quad (3)$$

where $F_{ij}^\alpha(\vec{n} + \vec{k}, \vec{n})$ indicates the number of commuter trips¹ (transitions) between the communities $i \rightarrow j$ for a member of subpopulation α . The explicit depend-

¹ If panel data are available on the commuter-decision behaviour of the different agents of the system (micro-level), the configurational transition rates can directly be calculated via $F_{ij}^\alpha(\vec{n} + \vec{k}, \vec{n}) = \sum_{l \in \Gamma_i} p_{ij}^{(l)}(\vec{n}, \vec{k}_l)$, where one has to sum up over all individual trips of all commuter from community i to community j . This procedure is however very extensive, because of the required immense data base (Courgeau 1985).

ence of the individual terms on \vec{n} indicate that all contributions related to a change of the population distribution $\vec{n} \rightarrow \vec{n} + \vec{k}$ have been summed up. In this way a summation of all such terms yields the total transition rate.

In the next step the transition rates have to be specified for the decision process to commute (Fischer et al. 1988; Weidlich and Haag 1988). If $n_i(t)$ persons are at time t in community i , the “probability to commute” to another community will be proportional $n_i(t)$. In this way the number of trips between i and j is given by

$$F_{ij}^\alpha(\vec{n} + \vec{k}, \vec{n}) = n_i(t) \cdot p_{ij}^\alpha(\vec{x}; t), \tag{4}$$

where $p_{ij}^\alpha(\vec{n}, \vec{x})$ is the individual transition rate from i to j for a member of the subpopulation α , $\vec{k} = \{0, \dots, 1, \dots, 0, \dots, (-1), \dots, 0, \dots\}$ and $F_{ij}^\alpha(\vec{n} + \vec{k}, \vec{n}) = 0$ for all other \vec{k} . Of course, this transition rate depends among others on the explicit spatial distribution of the population \vec{n} and specific characteristics \vec{x} of the communities, e.g. labour demand, labour supply, housing market, accessibility measures, specific location factors, services available for companies and households as well as leisure facilities (Domencich and McFadden 1975; Pumain and Saint-Julien 1989).

It has been tested, that mainly three sets of indicators are of importance for the transition rate to commute $p_{ij}^\alpha(\vec{x}; t)$:

- attractiveness indicators $u_i^\alpha(\vec{x}; t)$ of the particular community i for the subpopulation α , which depend across-the-board on the distribution of labour demand and supply of the communes. It is commonly known that individuals (commuters) compare the attractiveness of the communities with respect to certain characteristics such as working and housing conditions. The probability not to work at the place of home i instead to work in community j (place of work) increases with increasing differences $u_j^\alpha(\vec{x}; t) - u_i^\alpha(\vec{x}; t) > 0$ of attractiveness (Fechner 1877; Weber 1909; Fazio and Zanna 1981). Without any loss of generality the attractiveness can be scaled

$$\sum_{i=1}^L u_i(\vec{x}; t) = 0$$

- resistance function $g_{ij}^\alpha(t_{ij}, v_i^\alpha; t)$, representing the spatial interrelation (accessibility) of communities, depending on travel time t_{ij} , as well as regional shadow-costs $v_i^\alpha(\vec{x}; t)$. The resistance function is modelled via

$$g^\alpha(t_{ij}, v_i^\alpha; t) = \exp\left(\frac{-\beta^\alpha(t) t_{ij}}{1 + \gamma^\alpha(t) t_{ij}} - v_i^\alpha(\vec{x}; t)\right) \tag{5}$$

- with deterrence parameters $\beta^\alpha, \gamma^\alpha$ and shadow-costs $v_i^\alpha(\vec{x}; t)$. The shadow-costs $v_i^\alpha(\vec{x}; t)$ take into account the heterogeneity of the communities. Shadow

costs act as barriers and reduce the attractiveness of a region for commuters. By definition, the shadow-costs have to fulfil the constraint

$$\sum_{i=1}^L v_i(\bar{x}; t) = 0. \tag{6}$$

- a time-dependent scaling parameter $\varepsilon^\alpha(t)$ which correlates with the global mobility to commute.

This leads to the following commuter flow model (trip distribution):

$$\begin{aligned} F_{ij}^\alpha(\bar{n}, \bar{x}; t) &= n_i(t) p_{ij}^\alpha(\bar{x}, t) \\ &= n_i(t) \cdot \varepsilon^\alpha(t) \cdot g_{ij}^\alpha(t, v_i^\alpha; t) \cdot \exp(u_j^\alpha(\bar{x}; t) - u_i^\alpha(\bar{x}; t)) \end{aligned} \tag{7}$$

where $F_{ij}^\alpha(\bar{n}, \bar{x}; t)$ indicates the number of commuter trips (transitions) between the communities $i \rightarrow j$ for a member of subpopulation α .

The probability distribution $P(\bar{n}, t)$ contains a huge amount of information compared with the empirical information (data base). Therefore, a less comprehensive description in terms of mean values is adequate. The mean population number in community i at time t defined as

$$\bar{n}_i(t) = \sum_{\bar{n}} n_i P(\bar{n}, t). \tag{8}$$

It is possible to derive equations of motion for the mean values directly from the master equation. For this purpose the master equation is multiplied by \bar{n} from the left and summed up via all states \bar{n} . However, the resulting equations are not yet self-contained, since the determination of the right hand side (rhs) requires the knowledge of the probability distribution $P(\bar{n}, t)$. However, if one assumes that the probability distribution is a well behaved, sharply peaked uni-modal distribution quasi-closed approximate mean value equations can be derived:

$$\begin{aligned} \frac{d\bar{n}_i(t)}{dt} &= \sum_{\alpha=1}^A \sum_{j=1}^L F_{ji}^\alpha(\bar{n}, \bar{x}; t) - \sum_{\alpha=1}^A \sum_{j=1}^L F_{ij}^\alpha(\bar{n}, \bar{x}; t) \\ &= \sum_{\alpha=1}^A \sum_{j=1}^L \bar{n}_j(t) p_{ji}^\alpha(\bar{n}, \bar{x}; t) - \sum_{\alpha=1}^A \sum_{j=1}^L \bar{n}_i(t) p_{ij}^\alpha(\bar{n}, \bar{x}; t) \\ &= E_i^W(t) - E_i^H(t) = NC_i(t) \end{aligned} \tag{10}$$

Therefore the master equation provides the link between decisions to commute on the micro-level and the commuter flows on the macro-level. The dynamics of the mean population number $\bar{n}_i(t)$ of community i can be calculated on the basis of the commuter flows $F_{ij}^\alpha(\bar{n}, \bar{x}; t)$ and $F_{ji}^\alpha(\bar{n}, \bar{x}; t)$ between the different communities i and j . The first sum on the rhs (right-hand side) of Eq. 9 equals the employees at the place of home $E_i^H(t)$, the second sum equals the employees at the place of work $E_i^W(t)$. Therefore the dynamics of the population redistribution on the macro-level depends on the development of the net commuters $NC_i(t)$.

Long-term effects, e.g. structural development effects have to be considered as well. It is reasonable to assume that the attractiveness $u_i^\alpha(\bar{x}, t)$ and shadow-costs $v_i^\alpha(\bar{x}; t)$ of a community depend on a set of socio-economic variables \bar{x} (among other things also on the population distribution $n_i(t)$). Therefore, the impact of commuting on the population redistribution is of importance (Fig. 3). Depending on the initial conditions, such as the distribution of population at a given time and the further system parameters, the non-linear dynamics lead to self-organised commuter flow pattern (Nijkamp and Reggiani 1992; Goodwin 1994).

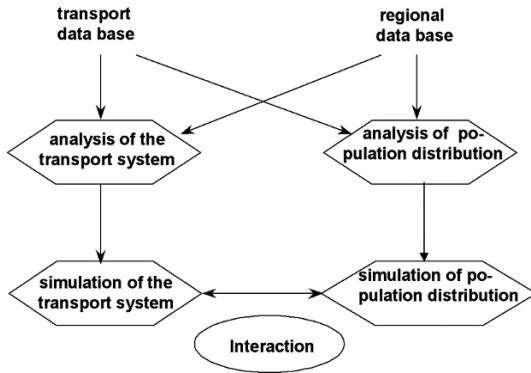


Fig. 3. Principle structure of the nested transport and urban model

3.2 Calibration of the Stuttgart Model

The system parameters, such as mobility $\varepsilon^\alpha(t)$, attractiveness $u_i^\alpha(\bar{x}, t)$, shadow-costs $v_i^\alpha(\bar{x}; t)$ and the resistance function parameters $\beta^\alpha(t)$ and $\gamma^\alpha(t)$ can directly be linked to the empirical (statistical registered) commuter flow matrices $F_{ij}^{\alpha emp}(t)$ (index *emp*), and the population numbers $n_i^{emp}(t)$, respectively. The minimisation of the functional (entropy-estimation)² (Wilson 1970, 1981)

² Using test series it was determined that the least-square-estimation represent the single flows $F_{ij}(t)$ much better then the entropy procedure estimation. On the other hand, the specific entropy-estimation takes into account that the origin and destination flows of each community are equal to the employees at the place of home (empirical origin flows) and employee at the place of work (empirical destination flows). This property of the entropy procedure can also be derived analytically by geometric programming (Kádas and Klafszky 1967).

$$\begin{aligned}
 G(u_i, v_i, \varepsilon, \beta^\alpha, \gamma^\alpha; t) &= \text{MIN} \left(\sum_{\alpha=1}^A \sum_{i,j=1}^L F_{ij}^{\alpha \text{ emp}}(t) \cdot \ln \left(\frac{F_{ij}^{\alpha \text{ emp}}(t)}{F_{ij}^\alpha(t)} \right) \right) \\
 &\approx \text{MIN} \left(\sum_{\alpha=1}^A \sum_{i,j=1}^L \left(\frac{F_{ij}^{\alpha \text{ emp}}(t) - F_{ij}^\alpha(t)}{F_{ij}^\alpha(t)} \right)^2 \right)
 \end{aligned}
 \tag{11}$$

with the constraint

$$\sum_{\alpha=1}^A \sum_{i,j=1}^L F_{ij}^{\alpha \text{ emp}}(t) = \sum_{\alpha=1}^A \sum_{i,j=1}^L F_{ij}^\alpha(t)
 \tag{12}$$

enables one to calculate an optimal set of system parameters $(u_i^\alpha(\bar{x}, t), v_i^\alpha(\bar{x}; t), \varepsilon^\alpha(t), \beta^\alpha(t), \gamma^\alpha(t))$.

In a second step, the estimated attractiveness and shadow-costs are linked to particular location factors (key-factors) $x_i^n, n = 1, \dots$ gained from two different fields: from a class of the so-called synergy variables, describing general group effects (positive and negative network externalities), and from a sequence of potential explanatory indicators, e.g. number of available jobs, the number of vacant dwellings, regional income per capita, shop distribution and other local infrastructure depending factors. The set of explanatory variables and its corresponding elasticity's are determined via a multiple regression analysis:

$$u_i^\alpha(\bar{n}, \bar{x}) = \sum_n a_n^\alpha x_i^n \qquad v_i^\alpha(\bar{n}, \bar{x}) = \sum_n b_n^\alpha x_i^n
 \tag{13}$$

The elasticity's a_n^α, b_n^α assigned to the socio-economic variables x_i^n are dimensionless numbers and indicate the influence of the independent variables on the dependent variable. The selection of relevant indicators is performed using appropriate statistical characteristics (T-values, other significance tests).

The results of the regression for the attractiveness (shadow costs) explains about 83% (75%) of the data variation (Haag and Binder 2001b). Because commuting describes trips from home to work and vice versa, the explaining variables are related to the following :

- Labour market (distribution of work places, wages,...)
- Housing market (distribution of apartments, dwellings, houses, price of land, rent level,...)
- Accessibility (travel time, travel costs, parking possibilities,...)
- Other specific indicators (availability of different services, neighbourhood, environment, recreation possibilities,...)

4 Results of the Simulated Scenarios

The following table details the simulated measures aiming to reduce the negative impacts of urban sprawl in the Stuttgart case city (SCATTER 2004). Of course, some of those applied scenarios are currently not realizable and are more of pure scientific interest.

Policy code	Description of the policy Stuttgart
0	Reference scenarios
001, 002, 003	Different reference scenarios: <ul style="list-style-type: none"> • 001S = without motorway A81, without S1 extension, without road tunnel Kappelberg • 002S = with motorway A81, with S1 extension, without road tunnel Kappelberg • 003S = with motorway A81, with S1 extension, with road tunnel Kappelberg <p>The reference scenarios (001S) are used for all policy codes 111S – 114S, reference scenario (002S) is used for policy code 115S, all other policy codes refer to the reference scenario (003S). The time horizons are:</p> <p>horizon (001S): 1995 horizon (002S): 2015 horizon (003S): 2020</p>
1	Transport infrastructures / services: radial infrastructures decreasing travel times between centre and periphery
11	<i>Implementation of a radial transport infrastructure linking centre and periphery: rail infrastructure, motorway, buses, HOV</i>
111 Common policy	111S: Extension of the light rail (S-Bahn) S1 (parallel to the corridor of the motorway A81) without motorway (length 16 km)
	This is tested on the 001S reference scenario, to which also the following measures (policy codes 112S, 113S, 114S) in the Stuttgart case city are compared.
112 Local policy	112S: Completion of the missing link of the motorway A81 in 1978 (length 23.9 km) , without S1 (light rail) parallel to the corridor of the A81
113 Local policy	113S: Completion of the missing link of the motorway A81 in 1978 (length 23.9 km), and extension of the S1 (light rail) parallel to the corridor of the A81 in 1992 (length 16 km)
114 Local policy	114S: Completion of the missing link of the motorway A81 in 1978 (length 23.9 km), and extension of the S1 (light rail) parallel to the corridor of the A81 in 1992 (length 16 km), with park&ride facilities (6 park&ride facilities, 7.500 new Parking spaces (about 19%))
115 Local policy	115S: 114S and building of a new road tunnel (tunnel Kappelberg) of the Bundesstrasse B29 in east-direction (Schwäbisch Gmünd)
12	<i>Implementation of a transport infrastructure with radial and tan-</i>

Policy code	Description of the policy Stuttgart
121 Common policy	<i>gential components (the latter one thus provides improved services for trips from periphery to periphery)</i>
2 211 Local policy	External factor: relocation of work places 211S: Relocation of 10.000 workplaces from Esslingen and Stuttgart-Untertürkheim to Sindelfingen (due to a shift of a production plant of DaimlerChrysler) tested on reference 003S
3 31 311 Common policy	Land use measures having an influence on urban sprawl <i>Fiscal measures applied to residential developments</i> 311S: Annual tax (development impact fee) applied on households locating in non-urban zones (about 670 EUR/household/year) and redistribution of the revenue of impact fee to the urban areas, as fiscal incentive to all households located in urban zones (Stuttgart, Ludwigsburg, Sindelfingen, Böblingen, Esslingen and Göppingen) tested on reference 003S
32 321 Common policy	<i>Regulatory measures applied to companies, inspired form the ABC theory</i> 321S: ABC-type policy applied to a part of the tertiary sector: <ul style="list-style-type: none"> • obligation (regulatory measure) for all jobs of the employment sector “business services”, to locate in A-type zone • an A zone is a zone of the capital of a district (NUTS3). In general those zones are also served by high quality public transport at regional scale. In these scenario, there are 7 A-zones in the Stuttgart Region tested on reference 003S
33 331 Common policy	<i>Fiscal measures applied to companies, inspired form the ABC theory</i> 331S: ABC-type policy applied to a part of the tertiary sector: <ul style="list-style-type: none"> • tax on new jobs of the employment sector “business services” locating in non-A-type zone; the tax amounts to 976 EUR/job • an A zone is a zone of the capital of a district (NUTS3). In general those zones are also served by high quality public transport at regional scale. In these scenario, there are 7 A-zones in the Stuttgart Region tested on reference 003S
4 41 411 Common policy	Measures aiming at a modal shift towards public transport by increasing travel costs or time by private car <i>Increase of car use cost</i> 411S: increase by 50 % of the cost per km for all drivers tested on reference 003S

Policy code	Description of the policy Stuttgart
412 Common policy	412S : cordon pricing (the cordon is located just inside the city of Stuttgart and the adjacent communes Ludwigsburg, Sindelfingen, Böblingen and Esslingen); tariff: 2,1 EUR/day applied to all drivers tested on reference 003S
5	Measures aiming at a modal shift towards public transport by decreasing travel costs or times by public transport, or by providing P&R facilities
51 512 Common policy	<i>Change in the fare of public transport</i> 512S : decrease of fare by 20%, applied to all public transport users tested on reference 003S
52 521 Local policy	<i>Park&ride facilities</i> 521S : Park and ride facilities see scenario 114S tested on reference scenario 113S
8	Combinations of selected measures
811 Common policy	Combination 811S = 411 + 512 + 311 <ul style="list-style-type: none"> • increase by 50% of the private car cost/km applied to all drivers • decrease of PT fare by 20%, applied to all public transport users • fiscal measure on residential developments: see scenario 311 tested on reference 003S
812 Common policy	Combination 812S = 411 + 512 + 331 <ul style="list-style-type: none"> • increase by 50% of the private car cost/km applied to all drivers • decrease of PT fare by 20% for home-work trips • ABC-type policy applied to a part of the tertiary sector: see scenario 331 tested on reference 003S
813 Common policy	Combination 813S = 411 + 511 + 311 + 331: <ul style="list-style-type: none"> • increase by 50% of the private car cost/km applied to all drivers • decrease of PT fare by 20% for home-work trips • fiscal measure on residential developments: see scenario 311 • ABC-type policy applied to a part of the tertiary sector: see scenario 331 tested on reference 003S

4.1 Simulation Results in the Stuttgart Case City

The following table presents the simulation results for the Stuttgart case city, with a set of basic indicators.

Table 1. Variation of the number of households in the urban zones (%)

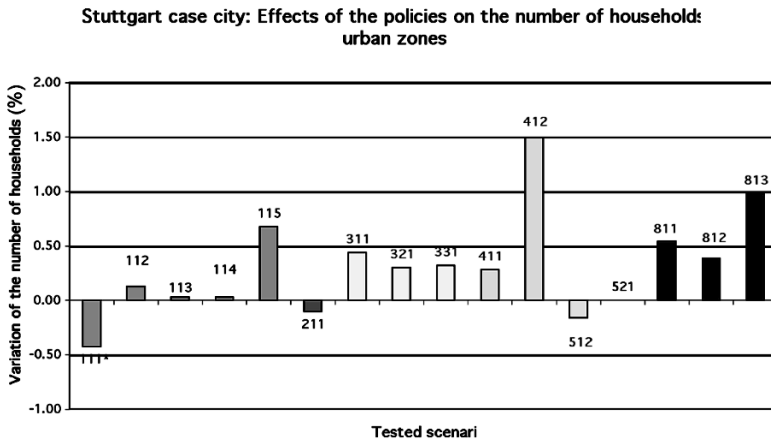
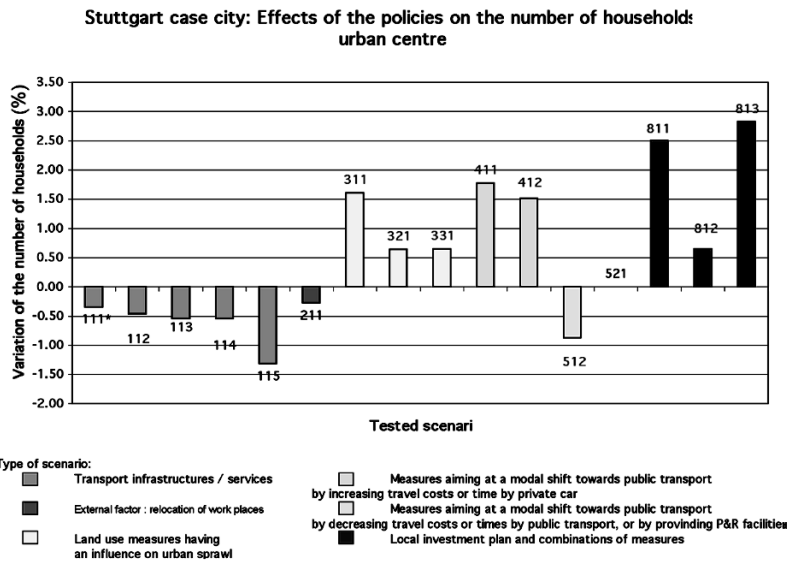


Table 2. Variation of the number of households in the urban centre (%)



*The effect of motorway A81 and light rail S1 (111, 112, 113, 114) is calculated by comparison with scenario 001 (situation without motorway and light rail). The effect of tunnel Kappelberg (115) is calculated by comparison with scenario 002 (which is also 114 – situation with motorway A81 and light rail S1). The effects of the other measures are calculated in comparison with scenario 003 (present state).

Table 3. Variation of the number of jobs in the urban zones (%)

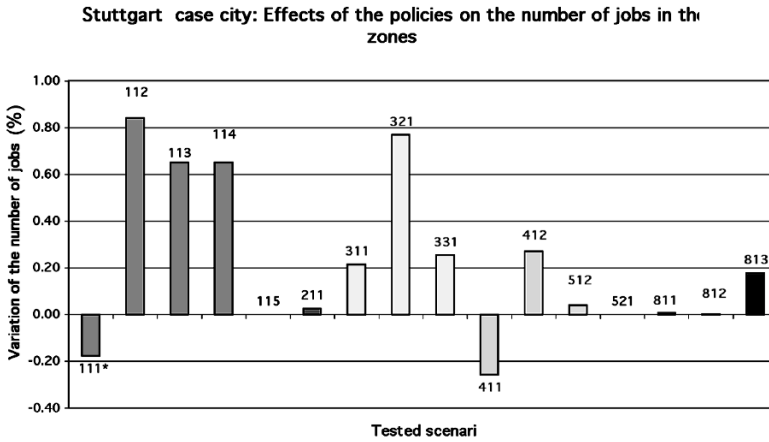
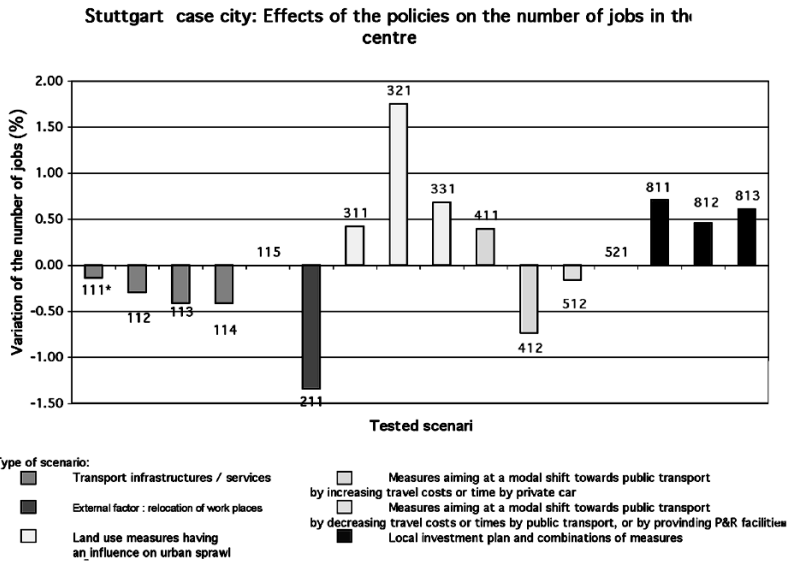


Table 4. Variation of the number of jobs in the urban centre (%)



*The effect of motorway A81 and light rail S1 (111, 112, 113, 114) is calculated by comparison with scenario 001 (situation without motorway and light rail). The effect of tunnel Kappelberg (115) is calculated by comparison with scenario 002 (which is also 114 – situation with motorway A81 and light rail S1). The effects of the other measures are calculated in comparison with scenario 003 (present state).

4.2 Examples of Two Tested Measures

As illustrative examples two measures are simulated and listed below:

4.2.1 Scenario 115S: Opening of the Kappelberg Road Tunnel

In this scenario (115S) the effect on urban sprawl a new road tunnel project (tunnel Kappelberg) in combination with the shift of the B29 is investigated (Figure 4). The measures are:

Completion of the missing link of the motorway A81 in 1978 (length 23.9 km), and extension of the S1 (light rail) parallel to the corridor of the A81 in 1992 (length 16 km), with park&ride facilities (6 park&ride facilities, 7.500 new Parking spaces (about 19%)) and building of a new road tunnel (tunnel Kappelberg) of the Bundesstrasse B29 in east-direction (Schwäbisch Gmünd)

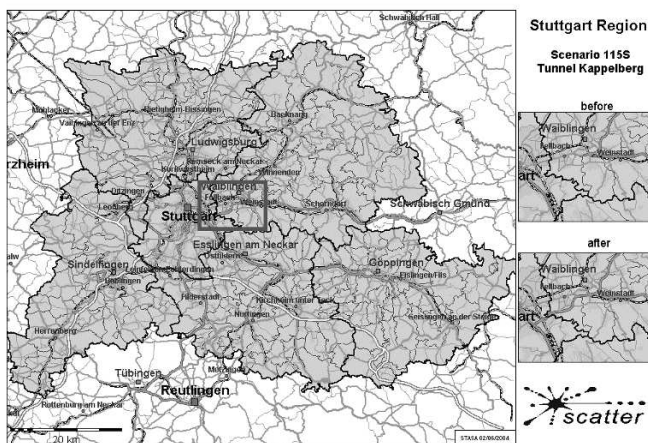


Fig. 4. Scenario 115S: Transport investment in radial direction – road tunnel Kappelberg (corridor Stuttgart – Aalen – Nürnberg)

The building of a new road tunnel (tunnel Kappelberg) of the Bundesstrasse B29 in east-direction (Schwäbisch Gmünd) also can be considered as an improvement of radial transport lines between the centre (Stuttgart) and more peripheral areas of the region.

According to the tables 1 – 4 it is obvious that this measure leads to a redistribution of households from urban centre towards urban zones between 1995 and 2015 (Fig. 5). An increase of total car mileages in the study area of about 2% and of the average home-work travel distance (about 1.4%), accompanied by a decrease of the public transport modal share (-0.6 points) is expected, since this measure favours car transport. The better link of the urban centre to the urban areas and hinterland leads also to an increase of the traffic speed and the negative effects of CO₂ emissions (+1.9%). The *H*-indicator (SCATTER 2004) states that an

increase of sprawl must be expected, but rather moderate compared with the effect of the extension of the motorway A81.

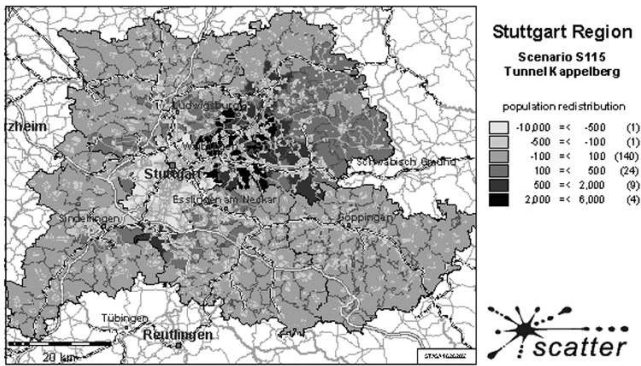


Fig. 5. Scenario 115S: Population redistribution

4.2.2 Scenario 211S: Relocation of Workplaces

In this scenario (211S) the impact of external factors such as the redistribution of workplaces from one part (east of city centre) of the Stuttgart region to the southern part, and its influence on the different socio-economic levels of the transport and settlement system is investigated (Fig. 6):

- **211S:** Relocation of 10.000 workplaces from Esslingen (5.000 workplaces) and Stuttgart-Untertürkheim (5.000 workplaces) to Sindelfingen (10.000 workplaces), due to a shift of a production plant of DaimlerChrysler.

It is assumed that in the short run the housing locations of the commuters are not changed (no constraint). In the long run, however, migration of households occur. Furthermore, it is assumed that the workplace distribution is fixed (with constraint).

The relocation of work places leads to changes of attractiveness values and shadow costs in the STASA- simulation model. This results in a redistribution of the commuter flows and population (housing locations). The different constraints take into account that all work places should be occupied and/or the housing population should be fixed.

For the region of Stuttgart a redistribution of work places is computed according to Fig. 6. This scenario seems to be rather realistic. Of course, the effect of this direct intervention on the distribution of jobs results in a huge change of the commuter pattern and effects the travel time relationships (travel time matrix) within a large area of the conurbation. In the short time consideration, it is assumed that all persons employed maintain their work place and their place of home.

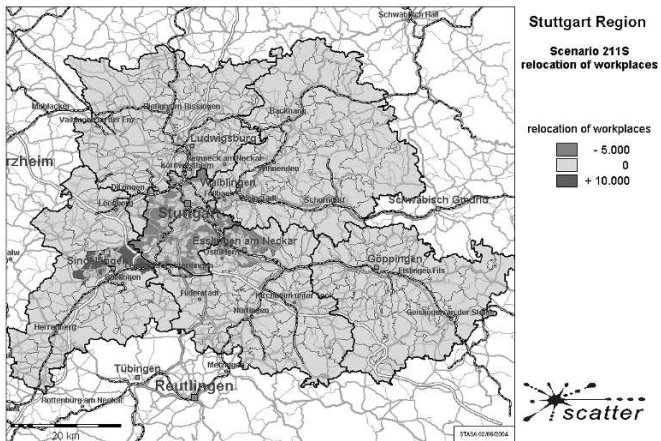


Fig. 6. Scenario 211S: Relocation of 10.000 workplaces within the area of the Stuttgart region

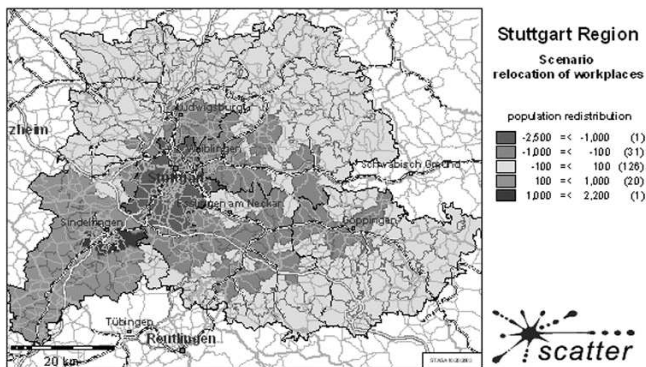


Fig. 7. Scenario 211: Population redistribution in the long run

In the long run, a partial redistribution of work places and housing locations (Fig. 7) may happen, resulting in a secondary redistribution of the commuter pattern. The quite different time scales of the different relocation processes (1 year change of work place, 10 years change of housing location) lead to a temporal separation of the effects. Therefore both scenarios are of importance in the city planning process. Because Sindelfingen is in the corridor Stuttgart – Singen located, the quite good accessibilities within this corridor may encourage households to search for a location towards Herrenberg with cheaper land-prices and therefore enhance urban sprawl. However, the distribution of the commuter flows in the study area before the shift of the considered production plant occurred (reference scenario 003S - horizon 2020) shows that this measure has rather positive

impacts related to the introduced indicators (Table 1 – 4). Due to a better spatial distribution of the commuter flows and the structure of traffic network, the total car mileages and the average travel times decrease slightly. Also the average home-work travel distance is rather not changed by this measure. The effects of this policy on the relative *H*-measure (SCATTER 2004) for inhabitants shows a very moderate increase, the sprawl on jobs, however, is quite considerable.

References

- Batty M, Besussi E, Chin N (2003) Traffic, Urban Growth and Suburban Sprawl. Paper presented at Mobilidade Para Todos, Automóvel Clube de Portugal, Lisboa, Portugal, 18 - 19 September 2003
- Binder J, Haag G, Koller M (2001) Modelling and simulation of migration, regional employment development and regional gross wage payment in Germany: The bottom-up approach. In Uddevalla Symposium 2001 – Regional economies in transition. University of Trollhättan/Uddevalla. Research reports 01, vol. 1, pp. 119-142
- Binder J, Haag G, Rabino G (2003) Analysis and modelling of commuter flows: Application to the regions of Stuttgart and Turin. *Jahrbuch der Regionalwissenschaften* **23**: 117-139
- BMVBW (1999) Modellierung agglomerationsbedingter Einflüsse auf wegebauinduzierte Beförderungsprozesse für die Bundesverkehrswegeplanung. Der Bundesminister für Verkehr, Bau- und Wohnungswesen, FE-Nr. 90546/1998
- Brown LA, Holmes J (1971) The Delimitation of Functional Regions, Nodal Regions and Hierarchy by Functional Distance Approaches, *Regional Science* **11**: 57-72
- Domencich TA, McFadden D (1975) *Urban Travel Demand: A Behavioural Analysis*. North Holland, Amsterdam
- EUROSIL 1999: European Strategic Intermodal Links, Project funded by DG VII, within the 4th Framework Programme
- Fazio RH, Zanna MP (1981) Direct Experience and Attitude-Behaviour Consistency. In: Berkovitz L (ed) *Advances in Experimental Social Psychology* **51**: 161-202
- Fechner GT (1877) *In Sachen der Psychophysik*. Breitkopf und Haertel, Leipzig
- Fischer M, Haag G, Sonis M, Weidlich W (1988) Account of Different Views in Dynamic Choice Processes. WSG Discussion Paper 1/88, Wirtschaftsuniversitaet, Wien
- Gayda S, Boon F, Schaillée N, Batty M, Besussi E, Chin N, Haag G, Martino A, Lautso K, Noël C, Dormois R (2003a) The SCATTER project- Sprawling cities and transport: From Evaluation to Recommendation. Paper presented at the European Transport Conference 2003, Strasbourg, October 2003
- Gayda S, Boon F, Schaillée N, Batty M, Besussi E, Chin N, Haag G, Martino A, Lautso K, Noël C, Dormois R (2003b) The SCATTER project- Sprawling cities and transport: From Evaluation to Recommendation. Paper presented at the 5th Biennial of Towns and Town Planners, Barcelona, April 2003
- Goodwin PB (1994) *Traffic Growth and the Dynamics of Sustainable Transport Policies*. ESRC Transport Studies unit, University of Oxford
- Haag G, Weidlich W (1984) A Stochastic Theory of Interregional Migration. *Geographical Analysis* **16**: 331-357

- Haag G (1989) *Dynamic Decision Theory. Applications to Urban and Regional Topics*. Kluwer Academic Publ., Dordrecht
- Haag G (2002) *Sprawling Cities in Germany*. FrancoAngeli, Milan
- Haag G, Binder J (2001a) *Vergleichende Untersuchung der Wirkungszusammenhänge zwischen Pendelverkehr und der räumlichen Organisationsstruktur der Region Stuttgart und der Provinz Turin. Study in order of the Stuttgart Region*. Steinbeis-Transfer Centre Applied System Analysis, Stuttgart
- Haag G, Binder J (2001b) *Modellgestützte Analyse zur regionalen Entwicklung von Beschäftigungsvolumen, Lohnsummen und Beitragseinnahmen*. Study in order of the Institute of Labour and Employment Research (IAB) Nuernberg. Steinbeis-Transfer Centre Applied System Analysis, Stuttgart
- Haag G, Gruetzmann K (2001) *Transport and Urban Development*. In: Clark G, Madden M (eds) *Regional Science in Business*. Springer, Berlin
- Haag G, Rabino G, Binder J (2001): *Vergleichende Untersuchung der Wirkungszusammenhänge zwischen Pendelverkehr und der räumlichen Organisationsstruktur einer Region/Metropole, Fallstudien in Italien und Deutschland*. DAAD-Project VIGONI, 314-vigoni-dr, Stuttgart
- Holmberg I, Johansson B, Strömquist U (2001) *A simultaneous model of long-term regional job and population changes*. Jönköping International Business School
- Johansson B, Klaesson J, Olsson M (2001) *Time Distances and Labour Market Integration*. In Uddevalla Symposium 2001 – *Regional economies in transition*. University of Trollhättan/Uddevalla. Research reports 01, vol 1, pp 255-279
- Kádas SA, Klafszky E (1967) *Estimation of the parameters in the gravity model for the trip distribution*. *Regional Science and Urban Economics* 6: 439-457
- Mills ES, Carlino G (1989) *Dynamic of county growth*. In: Andersson AE, Batten DF, Johansson B, Nijkamp P (eds) *Advances in spatial theory and dynamics*. North-Holland, Amsterdam, pp 195-206
- Nijkamp P, Reggiani A (1992) *Interaction, Evolution and Chaos in Space*. Springer, Berlin
- Oppenheim N (1995) *Urban travel demand modeling: From individual choices to general equilibrium*. John Wiley & Sons, London
- SCATTER (2004) *Sprawling Cities and Transport: from Evaluation to Recommendations*, European Commission, Key Action 4: *City of Tomorrow and Cultural Heritage*, EVK4-CT-2001-00063 RS
- Steierwald G, Kühne HD (eds) (1993) *Stadtverkehrsplanung*. Springer, Heidelberg
- TRANSECON (2004) *Urban Transport and local Socio-Economic Development*, European Commission, GMA1-2000-27049
- Weber A (1909) *Über den Standort der Industrien*. Mohr, Tübingen
- Weidlich W, Haag G (1984) *Concepts and Models of a Quantitative Sociology*. Springer Series in Synergetics. Springer, Berlin
- Weidlich W, Haag G (1988) *Interregional Migration*. Springer, Berlin
- Weidlich W, Haag G (eds) (1999) *ITEM: Integrated Transport and Evolutionary Model*. Sponsoring Project by DaimlerChrysler, Springer, Berlin
- Wilson AG (1970) *Entropy in Urban and Regional Modelling*. Pion, London
- Wilson AG (1981) *The evolution of spatial structure*. In: Bennett RJ (ed) *European Progress in Spatial Analysis*. Pion Books, London

Study of Urban Developers' Behavior in a Game Environment

Erez Hatna, Itzhak Benenson

Environmental Simulation Laboratory (ESLab), Tel Aviv University, Israel

Abstract

Most urban models accept the assumptions of the Markov processes: the state and location of each urban object at time step $t+1$ are defined by its state, location and environmental conditions at t . This assumption is not at all evident, just because the city is developed by humans who have memory and might implement long-term development plans, and, thus demands confirmations. The Markov nature of developers' behavior is investigated on the base of laboratory experiments, in which 30 participants were asked to construct a 'city' on the floor of a hall; each participant had to use the same set of mock-up buildings. Each mock-up established was represented as a feature of GIS layer, and its urban function, given by the participant, was recorded. The analysis of participants' behavior reveals that the relation between the urban pattern on the step t of the experiment and the decision regarding the urban function and location of the mock-up at $t+1$ is very close to assumed by the Markov theory. Based on the experimental results, spatially explicit model of participants' behavior was further constructed. The comparison between the experimental and the model patters, in their dynamics, clearly favor the idea of a shared Markov process as the basis for representing human urban development behavior. In the same time, with the increase in city complexity, the spectrum of participants' behavior becomes wider than that of the model and some participants tends to deviate from it. This experiment is a preliminary yet important step towards the experimental study of decision-making behavior among real developers and planners, which provide the basis for description of the real-world urban dynamics.

1 Introduction

Forty years ago, urban modeling took off the field of land-use dynamics up to the heaven of comprehensive description of urban systems' dynamics. The reality has melted the wasp thread of the thousands of relationships, nobody hoped to actually estimate. Now we are in the mid of the second twist of the spiral, which began in early 1990 with high-resolution Cellular Automata (CA) models and, in time, absorbed the fashionable Multi-Agent Systems (MAS). The foundations of MAS and CA have recently been combined and incorporated in the Geographic Automata

System (GAS) framework, which treats urban systems as collectives of geographic objects, fixed, like the infrastructure elements of CA, or non-fixed, similar to the mobile agents in MAS (Benenson and Torrens 2004).

Numerous GAS-models of land-use dynamics, car, and pedestrian traffic, residential and service migration implement an object-based perspective, in which the modeler formulates the rules of the object's changes in space-time, that is, cell transition rules in CA, rules of agent behavior in MAS or, more generally, object automation rules in GAS.

There are no "proves" that new generation of models works better than the previous one; however, the (very personal) "feelings" of the moment are that they do. The view of the models that make us feel so reveals a definite tendencies and fashions. First, these models are all on the solid ground of specificity – they separate one component of the urban system, not only traditional land uses, but also residential distribution, and, the hit of the last few years – traffic. Second, the rules of the urban objects' transitions or urban agents' decision-making are *simple*.

The simplicity of the agents' decision-making rules is especially important. Our imitation of human behaviors will always be over-primitive – why we could simulate the city, which is a collective outcome of these behaviors? The trade off between simplicity of the agent formalization and the ability to resemble their real-world behavior is crucial in complex human systems, just because the real world agents – human decision maker – are complex system themselves. In this paper, we investigate this problem regarding one of the basic properties of humans - their *memory*. We establish an experimental framework, which makes possible to investigate the importance of the decision-makers' memory when modeling land-use dynamics.

2 From "Memory" to "Delay:" Markov Process as a Landmark

In what follows, we limit ourselves to the quite simplistic discrete-time and probabilistic view of urban agents' behavior, and assume that

- *State of an agent X* at moment t is characterized by a vector X_t (include X 's location in space);
- *Behavior B* means simply the change of X_t ; $B : X_t \rightarrow X_{t+1}$
- *State of the environment E* at t - E_t - is the state of all the other agents and objects, an agent X accounts at t in order to behave;

At the same level of simplification, account for *memory* means introduction of *delay* into the description of the agent's behavior. Let us distinguish between three levels of accounting for memory/delay:

No memory: The state of an agent at $t+1$ does not depend on the agent or environment past, that is:

$$X_t = f(t); \quad (1)$$

Dependence on immediate past (Markov process): Each subsequent state of an agent is determined exclusively by its current state and the state of the environment. An object's automation rules can thus be presented as:

$$X_{t+1} = f(X_t, E_t) \quad (2)$$

Dependence on history: rules of the agents' behavior should explicitly include the object's as well as the system's history:

$$X_{t+1} = f(X_t, E_t, X_{t-1}, E_{t-1}, X_{t-2}, E_{t-2}, \dots) \quad (3)$$

Urban systems are essentially Markov and do not go beyond Markov; model (1) is basically sufficient;

It is hard to imagine the real-world agents that behave according to the predetermined rule (Eq. 1), and do not react to changes in their state or the state of the environment. That is why majority (if not all) urban CA and MAS models begin with the rules of type (2).

The exclusive dependence of a system's state at $t+1$ on its state at t is characteristic of the *Markov process*, introduced about a century ago by the Russian mathematician Andrey Markov. The basic Markov model is much more specific than Eq. 2:

- Objects can change their *states* at discrete moments of time,
- The set of states $S = \{s_1, s_2, s_3, \dots, s_k\}$ is assumed to be finite, and
- For each pair of states $\langle s_i, s_j \rangle$, there exists a *constant probability of transition* p_{ij} of agent's state from s_i to s_j ($s_i \rightarrow s_j$) during a time step (therefore, $\sum_j p_{ij} = 1$).

Let us note, that for a basic Markov model, given a matrix of transition probabilities $P = \|p_{ij}\|$, the probabilities that an agent will change its state from s_i to s_j not in one, but in two, three, ..., n , ... steps can thus be calculated as elements of the second, third, ..., n -th power of $P^2, P^3, \dots, P^n, \dots$ (Isaacson and Madsen 1976).

The basic result of a Markov process is that the *distribution of objects by states* converges to equilibrium over time, with equilibrium analytically calculated on the basis of matrix $P = \|p_{ij}\|$ (Isaacson and Madsen 1976). More complex (cyclic, for example) behavior demand that analytical relationships hold between the probabilities p_{ij} .

The basic Markov process is a very special case of Eq. 2; in particular, environmental influences (E_t) on transition probabilities p_{ij} are ignored. *Markov fields* or as otherwise called, *probabilistic Cellular Automata*, include local environmental effects into the model (Benenson and Torrens 2004). Each cell of probabilistic CA is explicitly located in space; its environment E_t is therefore defined by the objects *neighboring* on the cell. Transition probabilities p_{ij} thus depend on the cell state X_t and, in addition, on the states of the neighboring objects in E_t .

Extensions of the Markov field model allow for the variation of E_t with t , for accounts for non-local characteristics, such as the distance to the nearest cell of a specific (e.g., ‘road’) type, or for the global factors (e.g., average density of ‘dwelling’ cells over the model area) land-use (Yeh and Li 2002). However, the fundamental assumption of the Markov process — exclusive dependence of the object’s state at $t+1$ on its state at t — is taken for granted.

All land-use models we are aware of are applications of the Markov process; these consider urban areas as consisting of many parcels, each of which is in one of several discrete, easily recognized land-use states S — dwelling, industry, open land, and so forth. The probabilities of the parcel’s transition p_{ij} from state s_i into s_j are estimated by comparing land-use maps observed at consecutive moments of time (two or three land-use maps constructed for different years are usually available during an application). Investigation of CA land-use models focuses on future equilibrium distributions of land-uses and convergence to that equilibrium (Jahan 1986; Boerner et al. 1996; Gobin et al. 2002).

Why are we so sure of this formulation? Is there any proof that Eq. 2 is sufficient for urban modeling and that accounting for memory is superfluous?

Surprisingly, not much has been done toward testing which of the three assumptions (if any) best fits real-world situations. Model developers usually accept the Markov view (Eq. 2) as ‘reasonable’ and model the system in this way. The possibility (Eq. 3) of the real memory is simply ignored. Three reasons of this evasion can be posited (Benenson and Torrens 2004):

- Experimental data regarding the significance of agents’ memory or delays in objects’ dynamics are scarce;
- Theoretical demonstrations indicate that delays make model dynamics ‘too complex’;
- Most important - models without the delay seem to work well enough in practice.

A number of successful applications indirectly support Markov view regarding land-use in cities and agricultural areas and provide successful prediction for about 60-85% of the land-unit changes. The transition rules usually account for population density, major landscape factors (elevation, slope, and suitability of soil for agriculture) and accessibility measures (distance to the main river, the market, the main settlement, the water stream and the main road) (Gobin et al. 2002; LaGro and DeGloria 1992; Lopez et al. 2001; Brown et al. 2002; Hathout 2002). In social research, Markov applications include simulation of the residential distribution (Seetharaman 2003; Benenson et al. 2002).

Land-use models in which the sufficiency of the dependence of X_{t+1} on X_t and E_t alone was *tested* before the model was applied are few and provide support for each of the views (1) – (3). Markov view is supported by thorough investigations of Robinson (1978a, 1978b), who examined the land use change during 1962-1975 in three townships on the fringe of Akron, Ohio, and demonstrated that knowledge of the land use category at year $t - 1$ provides a significant decrease in the uncertainty of the land use type at year t , while the knowledge of X_{t-2} and X_{t-3} does not. Weng (2002) investigated the stability of land-use transition probabili-

ties for an area of rapid change across the Zhujiang Delta coastal region, and demonstrated insufficiency of Markov model - the values of p_{ij} , as well as their dependence on environmental factors differed significantly over three three-year time-spans. We are also aware of one study (Bell 1974), which revealed that knowledge of previous land-use was *not helpful* when predicting changes.

So, 'feelings' and practice of the model developers, remain in favor of the Markov view although empirical tests are scarce and ambiguous. However, the three views we formulated are not philosophical views of the world; they regard processes that can be measured and investigated. In particular, the choice between the models (1) – (3) can be justified if we:

1. Consider system that is fully controlled by its developer;
2. Construct the dependency of the developers' actions, performed at t and defining the situation at $t+1$ on their states and the state of the environment at $t, t-1, \dots, t-s, \dots$;
3. Find the minimal delay s in above description of the developers' behavior that is sufficient for reproducing the system's development.

In what follows, we perform such a test with a 'city game' in which 'developers' construct an artificial city by locating 'buildings' on the floor of a large room.

3 Description of the Experiment

30 undergraduate geography students were asked to construct a 'city' using 52 mock-up models on the trapeze floor of a hall 28.2 m² in area. The set of mock-ups was developed by Juval Portugali (Portugali 1996) and is used for various experiments (Portugali 2002). The mock-ups represent real buildings at a 1:100 scale; they also represent different urban functions — small- and medium-height dwellings, commercial sites, office building and so on. Average size of a mock-up is 20x20 cm, although it can range from 10x10 to 40x40 cm, and the overall capacity of the city area was about 700 mock-ups (Fig. 1).

Each participant constructed the 'city' only once by positioning mock-ups one at a time, in 52 *steps* of time t . During a time step, the participant chose a building from the remaining stock, declared its urban function and then positioned it. The participants were not permitted to relocate or return a mock-up to the original stock once it was positioned; the experiment was concluded once all 52 mock-ups were placed on the floor. When the experiments were carried out, the participant could declare one of seven urban functions for a building — Dwelling, Commerce, Entertainment, Culture, Religion, Industry, or Office. After analyzing the experiments, we combined five of the functions — Commerce, Entertainment, Culture, Religion, and Office — into one, Service.

At each step, the mock-up's urban function, identifier, position, and orientation were recorded and represented as features of a GIS layer (Fig. 2).

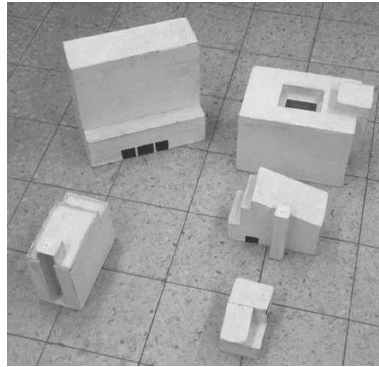


Fig. 1. 7 of the 52 mock-ups used in the experiments (the floor tiles are of 20x20cm size)

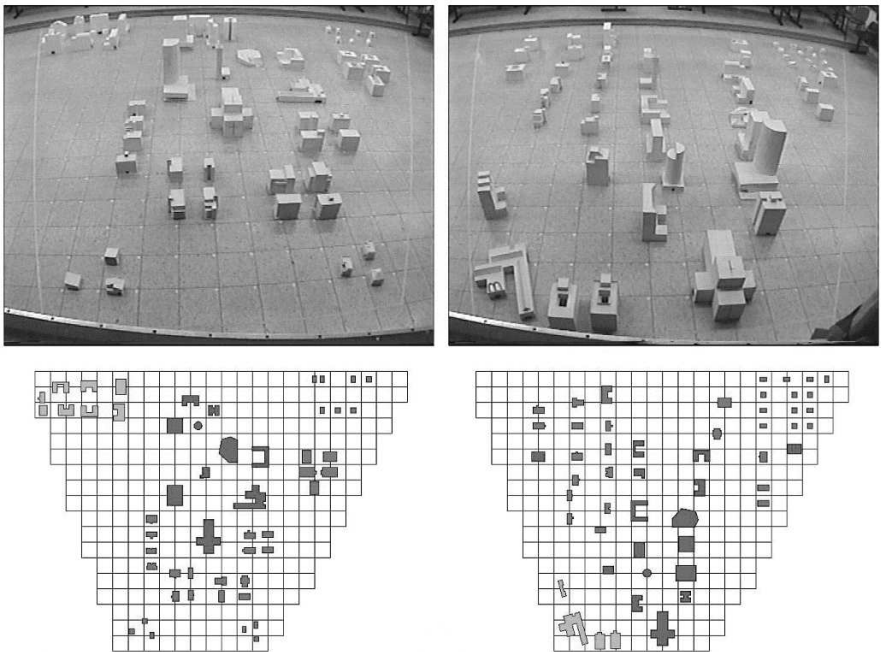


Fig. 2. A snap-shot of two games outcome (above), and their GIS presentation by means of mock-up foundation polygons (below).

Figure 3 presents the final patterns of 6 cities, randomly selected from the 30 available game outcomes.

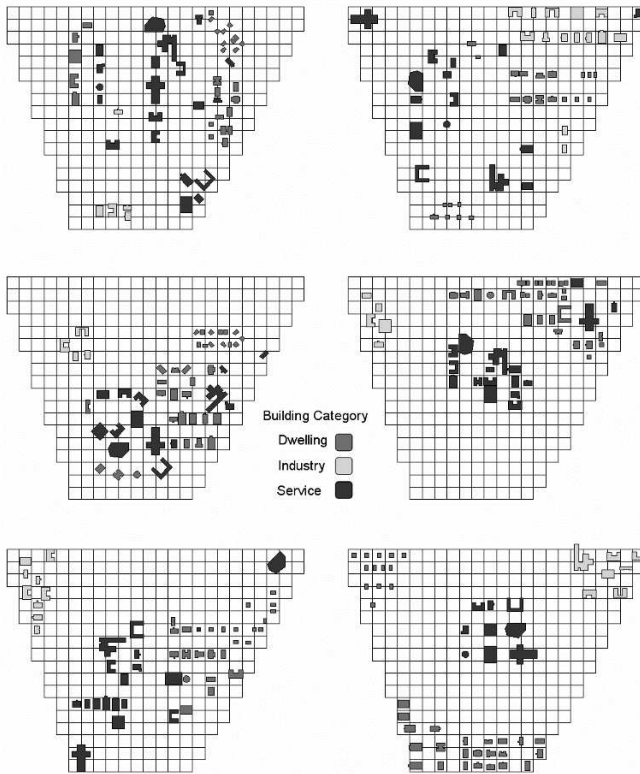


Fig. 3. Six typical cities built in the experiments

The basic problem we address regards the delay (s) necessary for adequate representation of the behavior of every participant. However, even if the behavior of all of them can be described by the Markov process ($s = 1$), the description itself can differ between the participants. 52 Behavioral acts are evidently insufficient for distinguishing whether two participants follow different (while, nonetheless, Markov) models, and in what follows we go half-way: first, assume that there exist a common model of participants' behavior and estimate its parameters, and, second, investigate individual deviations from this model¹. As stated in the introduction, our goal is to find minimal delay s , which is sufficient for the description of behavior of *all* participants.

¹ There are no principal difficulties in constructing personal model of participants' behavior – it simply demands several hundred behavioral acts to be executed, i.e. several hundred mock-ups available.

4 Analysis of the Experiment

Let $pattern_t$ denotes the urban pattern developed until reaching t ($t > 0$), and $actions_t$, $t = 0 \div 51$, the set of two elementary behavioral acts of the participant — choice of the mock-up’s function F_t , $F_t \in \{Dwelling, Service, Industry\}$ and establishing its position L_t . Let us consider each separately.

$Actions_0$ begin the game.

According to the models introduced in section 2, the participant’ behavior can be:

- *Simpler than Markov*, when $actions_t$ are independent of $pattern_{t-1}$ and $s = 0$;
- *Markov*, when $actions_t$ are determined by the $pattern_{t-1}$ and $s = 1$;
- *More complex than Markov*, when $pattern_{t-1}$ is insufficient for determining $actions_t$, and $s > 1$.

Let us note that the latter can indicate that the participant follows some pre-determined plan, say, first constructs industrial building, then completes dwellings, etc.

As mentioned above, we assumed that all participants shared the same model of behavior, and the data is considered as 30 repetitions of this model. We further assume that the choice of the building function is statistically independent of the choice of building location, and these two components of the participant’s behavior can be analyzed separately.

4.1 Choice of a Building’s Urban Function F

To verify the value of delay in building function choice we have to analyze the sequences of the choices. Let us begin with the $F_{t-1} \rightarrow F_t$ pairs of consecutive choices. Table 1 presents the experimental frequencies of observed transitions and the frequencies expected in cases of a choice of F_t independent of the choice of F_{t-1} :

Table 1. Actual frequencies of building function choice F_t , depending on function F_{t-1} of the building selected at the previous time step and expected frequencies when choice F_t is independent of F_{t-1} (in brackets)

F_{t-1}	F_t	Dwelling	Industry	Service
Dwelling		743 (570.8)	54 (115.1)	135 (246.1)
Industry		52 (115.1)	93 (23.2)	43 (49.6)
Service		142 (251.1)	42 (50.6)	22 (108.3)

The data in Table 1 are strongly in favor of the dependence of F_t on F_{t-1} ($\chi^2 = 556.6$, $p < 0.001$), with contingency coefficient $C = 0.52$. We can thus reject the simpler-than-Markov option and continue the analysis, assuming that $s \geq 1$.

Table 2 presents estimates of $F_{t-1} \rightarrow F_t$ transition probabilities in the case of Markov behavior, $s = 1$:

Table 2. Probability of choosing a building of a given function, given the building function chosen at the previous time step

F_{t-1}	F_t	Dwelling	Industry	Service
	Dwelling	0.797	0.058	0.145
	Industry	0.277	0.495	0.229
	Service	0.346	0.102	0.551

To test the *sufficiency* of the Markov approach regarding choice of building function, we tested whether the function of F_t depended on the function of a building selected *prior to* F_{t-1} . Markov theory suggests how this can be done — if the function choice process is Markov, then the probability that a building's function F_t is chosen given certain F_{t-s} is obtained from the matrix, which is power s of matrix P , obtained for $F_{t-1} \rightarrow F_t$ transitions (Table 2). We should thus compare the experimental frequencies of $F_{t-s} \rightarrow F_t$ transitions to the theoretical ones, calculated as powers P^s for $s > 1$. Table 3 presents the results of this comparison for $s = 2, 3$, and 4:

Table 3. Comparison of experimental and theoretical frequencies of $F_{t-s} \rightarrow F_t$ transitions for $s = 2, 3, 4$.

Time-delay s	χ^2	df	Significance	Contingency coefficient
2	6.169	8	p ~ 0.628	0.064
3	23.743	8	p ~ 0.003	0.126
4	23.459	8	p ~ 0.003	0.127

Table 3 confirms the Markov approach to participants' choice of building function regarding $F_{t-2} \rightarrow F_t$ transitions although the result does not hold at the same degree for delays $s > 2$, that is, $F_{t-3} \rightarrow F_t$ and $F_{t-4} \rightarrow F_t$ transitions. A review of the χ^2 components demonstrates that *Industry* \rightarrow *Industry* transitions are responsible for the discrepancy (Table 4).

Looking closer at the *Industry* _{$t-s$} \rightarrow *Industry* _{t} transitions, the observed and expected numbers of transitions are 60 versus 52.3, respectively, for $s = 2$, 50 versus 35.0 for $s = 3$, and 46 versus 27.5 for $s = 4$. That is, within the confines of the experiment, participants returned to *Industry* mock-ups a few steps later at a rate *more frequent* than might be expected assuming that choice of a building's urban function is indeed Markov. We did not proceed with this analysis because, given the data and the χ^2 statistics, we could conclude that the participants' choice of building function was not fully Markov but close to it.

Table 4. Components of χ^2 when comparing experimental and theoretical frequencies of transitions for time delays $s = 2, 3$, and 4.

B_{t-s} type	Transition B_t type	(Observed - Expected) ² /Expected		
		$s = 2$	$s = 3$	$s = 4$
Dwelling	Dwelling	0.890	3.711	2.354
Industry	Dwelling	0.494	3.516	1.864
Service	Dwelling	0.678	1.212	0.063
Dwelling	Industry	1.243	2.546	4.292
Industry	Industry	1.134	6.429	12.445
Service	Industry	0.314	0.330	0.826
Dwelling	Service	0.998	4.566	1.081
Industry	Service	0.038	0.202	0.472
Service	Service	0.382	1.232	0.062
Total χ^2		6.169	23.743	23.459

4.2 Choice of Building Location L

Participants' tendency to cluster buildings of the same function is readily observable to the naked eye (Fig. 3), but to test the hypothesis that a building's location at t depends on patterns established before, we have to specify the meaning of this dependency. For example, one might assume that L_t is determined by the position L_{t-1} and function F_{t-1} of the previously established building. We investigated several hypotheses of this kind and found that location patterns based on unitary buildings did not engender any results. What did work was the dependence of location L_t chosen for a new building of the functional type F_t on: (a) three distances between L_t and the nearest buildings of each of the three functional types observed in $pattern_{t-1}$ and (b) the direction of the vector that connects the nearest building of the functional type F_t in $pattern_{t-1}$, and L_t .

4.2.1 Dependence of L_t on the distances to the nearest buildings in $pattern_{t-1}$

Let us denote the locations of the Dwelling, Service, and Industry buildings nearest to L_t in $pattern_{t-1}$ as $nDwelling_{t-1}$, $nService_{t-1}$, $nIndustry_{t-1}$. The distances between L_t and $nDwelling_{t-1}$, $nService_{t-1}$, $nIndustry_{t-1}$ were estimated with the experimenters' GIS maps (Fig. 2, below). To illustrate, let F_t be the *Industry* mock-up located at L_t ; the nearest *Dwelling*, *Industry*, and *Service* buildings in $pattern_{t-1}$ are then found and the three distances between them and L_t are estimated. This analysis resulted in nine distributions (Fig. 4).

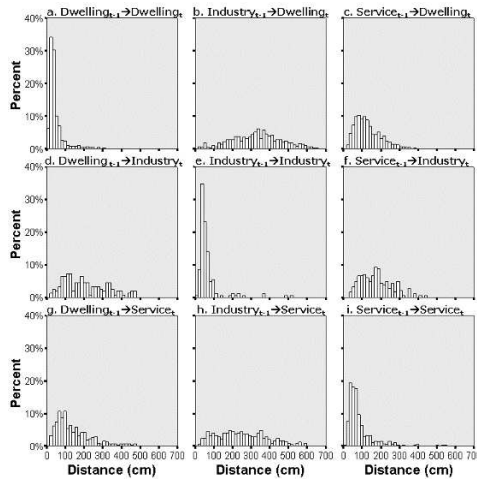


Fig. 4. Distributions of distances between nearest neighbors for 52 steps

As expected, the resulting distributions of distances reflect the participants' clustering tendency, and allows to reject the hypothesis that L_t does not depend on $pattern_{t-1}$. Indeed, according to Figs. 4a, 4e, and 4i, participants prefer to locate a new building near a standing building of the same functional type, while staying far away from the buildings of other types. The effect was strongest in the case of dwellings (Fig. 4a), while the distances between Industry and Dwelling buildings (Figs. 4b, 4d) indicated the greatest aversion between urban functions.

Location L_t also depended on the angle between the buildings. This dependence reflected the participants tendency to plan streets.

4.2.2 The Choice of L_t Reflects the Tendency to Plan Streets

Let us consider $pattern_t$, and denote as nL_{t-1} the location of the nearest to L_t , and as nnL_{t-1} the location of the nearest to nL_{t-1} buildings with function F_t (Fig. 5).

The analysis of the experiments' GIS maps reveals that the direction of the vector connecting nL_{t-1} to L_t , given by the angle ϕ_{t-1} in Fig. 5, depends on direction of the vector connecting nnL_{t-1} to nL_{t-1} (i.e. angle $n\phi_{t-1}$ in Fig. 5). As shown in Fig. 6, participants prefer to orient houses either along the line of already established buildings of the same functional type or perpendicular to that line. This dependence decreases when the distances between nnL_{t-1} , nL_{t-1} , and L_t grow; in what follow we employ for the distance below 70cm.

We did not proceed with the statistical tests of the dependencies between L_t and patterns $pattern_{t-s}$, where $s > 1$. Instead, we attempted to describe the participants' patterns based on the minimally necessary – Markov – representation of the process, in which we employ $s = 1$.

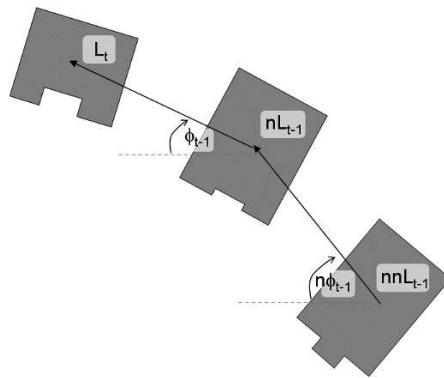


Fig. 5. Vectors (nnL_{t-1}, nL_{t-1}) and (nL_{t-1}, L_t) and the difference in their direction

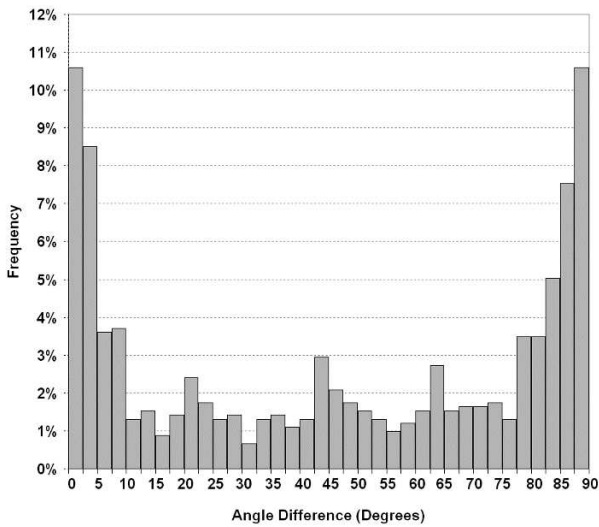


Fig. 6. Histogram of the difference in direction of the vectors connecting closest pairs of buildings of the same type at a distance below 70 cm (angle $\phi_{t-1} - n\phi_{t-1}$, according to the Fig. 5)

5 From Transition Rules to Integral Description of Participants' Behavior

Distance to the nearest neighbor and directions of the street network sprawl: Are these two factors sufficient to describe the behavior of the participant in our experiments? Are both taken into account by participants when they make a location decision?

To answer these questions, we constructed the Markov model of participant's behavior. The main assumption of the model is that the participant reacts on the distances to nearest neighbors and on the direction to the previously established building of the same type. Formally, we assume that the distributions in Fig. 4 and 6 and in Table 4 express these reactions, and consider these distributions as factor potentials: a *distance potential* regarding nearest neighbor of each of three types and a *direction potential* for expanding the street network. Assume that these potentials determine participants' location decisions; we can now investigate their sufficiency for reproducing the patterns that emerged in the experiments (Fig. 2). Before attempting to do so, we have to establish how participants combine these potentials.

Data on 30 artificial cities are sufficient to verify how participants combine *pairs* of the four potentials: The tests do confirm that the dependence of L_t on the distance to the *two* nearest neighbors in $pattern_{t-1}$ can be presented as a *product* of the marginal dependence of L_t on the distance to each member of the pair. To illustrate, Table 6 presents the experimental potential of establishing *Industry* when depending on the distance to the nearest *Industry* and *Service* buildings in $pattern_{t-1}$ versus the product of the potential for establishing *Industry* when depending on the distance to the nearest *Industry* and *Service* *separately* in the same pattern. Two calculation methods provide very similar results (Table 5):

Table 5. Observed and expected (in brackets) numbers of *Industry* buildings, depending on distances from previously located nearest *Industry* and *Service* buildings

		Distance from <i>Service</i> (cm)				
		0 – 60	60 – 120	120 – 180	180 – 240	Above 240
Distance from <i>Industry</i> (cm)	0 - 60	4 (4.5)	30 (29.2)	27 (26.0)	23 (24)	18 (18.2)
	60 - 120	3 (1.9)	10 (12.0)	9 (10.7)	11 (9.9)	9 (7.5)
	Above 120	0 (0.6)	5 (3.7)	4 (3.3)	3 (3.1)	1 (2.3)

Table 6 presents the χ^2 comparisons between several potentials, estimated according to the distances to the two nearest neighbors in $pattern_{t-1}$, and as a product of two one-dimensional potentials in 6 of 24 arbitrarily chosen cases:

Table 6. The results of the χ^2 -comparison between directly estimated two-dimensional potentials and the product of two one-dimensional potentials

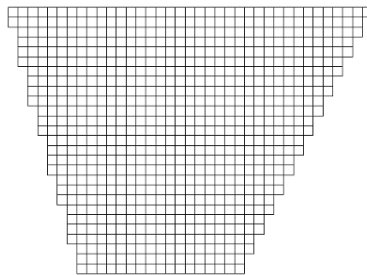
F_i -type	Nearest neighbor type	Nearest neighbor type	χ^2	df	Significance
Industry	Industry	Service	3.802	8	$p \sim 0.87$
Industry	Industry	Dwelling	11.831	10	$p \sim 0.30$
Service	Industry	Service	7.391	12	$p \sim 0.83$
Service	Industry	Dwelling	33.122	15	$p \sim 0.05$
Dwelling	Industry	Service	53.228	18	$p < 0.001$
Dwelling	Industry	Dwelling	27.436	18	$p \sim 0.07$

Supported by the two-dimensional test results, we assumed that participants follow the *product of all four potentials*.

6 The Simulation Model of Participants' Behavior

Our next step was simulating experimental cities by applying the above rules. Each 'developer' in the model built the city in 52 steps on a grid identical to the experimental grid. At each step t , the participant first assigns urban function F_t to the building according to the Markov transition matrix P given in Table 2, and then chooses its location L_t , based on the product of the distance and direction potentials for F_t with respect to $pattern_{t-1}$. We represent the city space as a trapezoid of cells, each 20x20cm in size, the same as the average size of a mock-up foundation (Fig. 7).

Each cell can contain one building only.

**Fig. 7.** Representation of the urban 20x20cm cell-space used in the model.

7.1 Modeling Building's Function F_t

We initiate each model run according to experimental estimates of the start frequencies for building of each of three types (Table 7):

Table 7. Probability to start the simulation with a building of a given type

Building function	Probability to choose at $t = 0$
Dwelling	0.300
Industry	0.067
Service	0.633

For $t > 0$, building type is assigned according to the basic Markov model, with the matrix of transition probabilities given in Table 2 (above).

7.2 Modeling Building Location L_t

We assume that the center of the trapezoid is a reference point for locating the first building in the city. Depending on the functional type, which is already chosen according to Table 7, the first building is located according to the distance to trapezoid center estimated on the base of 30 first steps (Table 8):

Table 8. Distributions of distances between buildings of a given function and the trapezoid center at $t = 0$

Distance from the trapezoid center (cm)	Probability to locate first building at the given distance		
	Dwelling	Industry	Service
0 – 25	0.00	0.00	0.50
25 – 75	0.22	0.00	0.30
75 – 175	0.22	0.00	0.05
175 – 275	0.33	0.50	0.05
above 275	0.22	0.50	0.10

Note that according to Tables 7 and 8, more than half of the buildings in the first step are *Services*, located close to the trapezoid's center.

In the model, we assumed that in order to locate F_t , a participant evaluates three distance potentials (as given in Fig. 4) and one direction potential (as given in Fig. 6) for all unoccupied at $t - 1$ locations, and chooses L_t according to their product. The potentials in Figs. 4 and 6 could be biased because of the high density of buildings in the latter time-steps. In order to reduce the bias in the estimates, we based the model on similar distributions obtained with the first 20 steps of each experiment; these distributions were very close to those presented in Figs. 4 and 6.

The algorithm of building locating is as follows. Let $V_t = \{v_{t,i}\}$ be the set of all empty cells at step t , and the function type of a building selected for locating at t is F_t . For each $v_{t,i} \in V_t$, three distances $d_{t,i}$ to the closest *Dwelling*, *Industry* and *Serv-*

ice buildings were then calculated and three distance potentials for F_t were estimated according to the distance potentials (as given in Figure 4, but reduced to the first 20 time-steps). To estimate the direction potential for each $v_{t,i}$, the locations nL_{t-1} of the nearest to $v_{t,i}$ and nnL_{t-1} of the nearest to nL_{t-1} buildings of an F_t -function were found and the distances between them were then calculated. If both distances were below 70cm, the direction potential of the $v_{t,i}$ was calculated according to the distribution presented in Fig. 6; otherwise, the direction potential of $v_{t,i}$ was set to uniform. To cope with the discreteness of the space, especially for close locations, we aggregated the frequencies in Fig. 6 according to the distance d between $v_{t,i}$ and the neighboring cell. For a given d , the potentials of the $N(d)$ possible directions (where $N(d)$ is the number of cells at a distance d from $v_{t,i}$) were calculated as a sum of the class frequencies (as presented in Figure 6, but reduced to the first 20 time-steps) over an interval of $\alpha(d) = 360^\circ/N(d)$ width.

The product of the three distance and one direction potentials for $v_{t,i}$ were then multiplied; the result was taken as the overall potential $a_{t,i}$ of cell $v_{t,i}$. The values $a_{t,i}$ were then normalized over V_t and the probability $p_{t,i}$ to locate a building of F_t -functional type in $v_{t,i}$ was obtained:

$$p_{t,i} = a_{t,i} / \sum_i a_{t,i} \quad (4)$$

Finally, the building was located at $v_{t,i}$ with the probability $p_{t,i}$.

8 Evaluation of Model Results

To evaluate the model we generated 500 patterns of 52 buildings and constructed the distributions of the distances between each building and its nearest neighbor for each functional type. Fig. 8 presents typical examples of the simulated urban patterns.

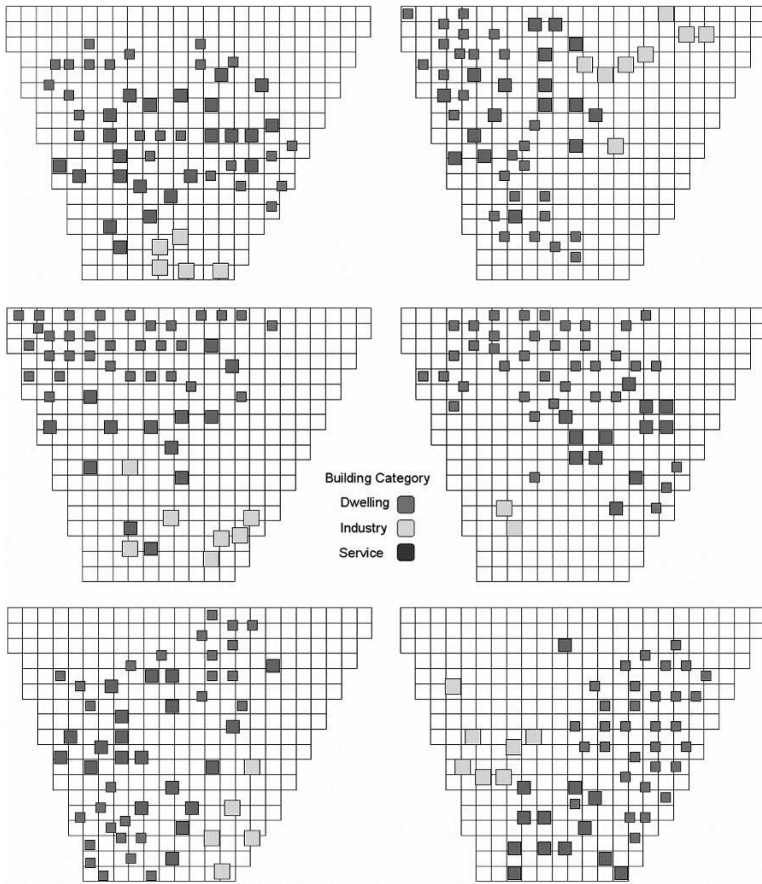


Fig. 8. Examples of the model patterns at step 52

As could be expected, the patterns expressed strong aggregation — buildings of the same function tended to cluster together, buildings of different functions were positioned further apart, industrial buildings tended to be located at the periphery, dwelling and service clusters were close together and so forth.

The means and standard deviations for the model and for the experiments are presented in Table 9. In all cases excluding one, the difference between the experimental and model means is less than 20% of the model mean; in half the cases the differences are insignificant. The largest difference is observed for the *Dwelling* → *Dwelling* pair: some of the difference can be explained by biased measurement of the distance between two close dwelling buildings. The model STD values are equal or lower than the corresponding experimental ones, while the maximal difference in the latter case is less than 30% of the model one. We therefore conclude that the basic correspondence between the model and reality is considerable.

Table 9. Means and standard deviations of the distance between building and its nearest neighbors in the simulated and experimental cities.

Pair	Nearest neighbors distance			
	Experiment		Model	
	Mean	STD	Mean	STD
$Dwelling_{t-1} \rightarrow Dwelling_t$	30	6	43	4
$Industry_{t-1} \rightarrow Industry_t$	58 ^a	32 ^b	65	37
$Service_{t-1} \rightarrow Service_t$	61 ^a	14 ^b	62	15
$Industry_{t-1} \rightarrow Dwelling_t$	280	97	258	59
$Dwelling_{t-1} \rightarrow Industry_t$	184	97	153	64
$Service_{t-1} \rightarrow Dwelling_t$	104 ^a	41	106	24
$Dwelling_{t-1} \rightarrow Service_t$	112	56	85	19
$Service_{t-1} \rightarrow Industry_t$	131 ^a	63	120	46
$Industry_{t-1} \rightarrow Service_t$	212 ^a	63 ^b	196	58

^aModel and experimental means do not differ significantly at $p = 0.95$.

^bModel and experimental STD do not differ significantly at $p = 0.95$.

We should stress that the experimental information included into the model does not automatically entail successful simulation. Indeed, the model incorporates important assumptions that simply *cannot be tested*:

- participants' decisions at each time-step are two-staged decisions, entailing choice of function and location;
- participants' location decisions regarding distance are determined by the distance to the building's nearest neighbors.
- participants' location decisions regarding direction are determined by the relative location of buildings of the same functional type only.
- In the course of decision-making, model developers adjust their behavior to the product of all four potentials
- All participants share the same averaged behavior, that is, substitution of individual participants has no effect on average participant behavior.

Although it is possible to investigate a spectrum of 'what-if' questions regarding the strength of one or more of the above assumptions, in the following we focus on the last assumption only and investigate individual participant's deviations from the "average" model. Regarding the other assumptions, the series of model runs indicates that the assumption that the participant selects the location of the next building according to the product of all four potentials is worth mentioning. Contrary to this assumption, in a series of experiments we assumed instead that the overall location potential is estimated according to the product of the highest distance and direction potentials obtained. This 'bounded rationality' view was, actually, our theoretical starting point with respect to the ability of human participants to combine potentials. However, the model patterns obtained under these assumptions corresponded much less to the experimental results than those based on the product of all the potentials, presented here.

The commonsensical conceptualization of the correspondence of the proposed Markov model with the experimental behavior is that whether participants remember their city development plans or not, is unimportant — the emerging patterns 'externally' represent all the information they need to proceed. Moreover, if the spectrum of the model patterns (which, by definition, reproduce the same average logic of development) covers the spectrum of experimental patterns, we can ignore possible differences in individual human reactions to the emerging pattern. Thus, are individual human participants nothing more than variations of the same behavioral archetype? This question deserves detailed investigation.

9 Behavior of the Human Participant Versus the Model

Despite good average fit, marginal disagreements between the model and experimental city patterns do remain. First, the test of the hypothesis regarding the Markov choice of the mock-up functional type resulted in significant χ^2 -values for three- and four-step time delays (Table 3). Second, some disagreement remains between the mean and variance of the nearest-neighbor distances, especially in the case of pairs of different types (Table 9); in this case, the model patterns revealed less variety than that obtained in the experiments.

We attempted to investigate those differences by altering the assumption that simulating the participants' 'average' behavior is sufficient; possible reason for the differences may rest on individual deviations from that 'average'. One can think of 'non-average' Markov participants' who consistently employ probabilities of building function choice differing from those presented in Table 2, and potentials functions different from those shown in Figs. 4 and 6, to say nothing about non-Markov participants who do not abide by any of our assumptions.

To identify possible individual deviations, we compared characteristics of experimental and simulated behavior not yet exploited — that is, we investigated the *potential of the locations selected by human participants at each time-step*. The idea captured here is that if the participant acts according to the 'average' Markov model at each time-step, then the potentials of the actually selected locations are sufficiently close to the highest possible at each time-step.

The correspondence between participants' location decisions and the model potentials was measured for each time step t ($t > 1$) of every experiment, by the fraction q_t of empty locations which have higher potential than the potential of the location actually selected by the participants. A low q_t value indicates a Markov type decision while a high value indicates an unlikely Markov decision. The average of q_t values $\langle q_t \rangle$ is used to describe the decisions strategies of each participant ranging from low $\langle q_t \rangle$ values of participants following Markov strategies, to high $\langle q_t \rangle$ values of participants with high deviation from the Markov model.

The values of $\langle q_t \rangle$ for each of the 30 experiments are presented in Fig. 9. The fractions are ordered from low to high values and seems to increase in a linear rate which changes, most notably, at $\langle q_t \rangle = 15\%$ and at $\langle q_t \rangle = 25\%$. The final patterns for selected experiments are also presented in Fig. 9. An overview of the

urban patterns give the impressions that almost all the patterns have some internal consistency in the method of their construction, while the low $\langle q_i \rangle$ experiments are simpler the higher valued experiments.

The final pattern produced by low $\langle q_i \rangle$ (Markov) participants is characterized, as expected, by a high degree of zoning and a few large clusters (Fig. 9, ranks 1,3,5,11). Patterns of higher $\langle q_i \rangle$ are characterized by a large number of clusters and high degree of zoning (rank 13, 18) or by a relatively high mixture of urban functions (rank 20, 22). Both groups deviated from the basic Markov model, the first maintained the high zoning tendencies but initiated 'too many' new urban districts (clusters). This may indicate participant with some extra creativity not found in the basic Markov model. The second group, which did not maintain a high zoned pattern, may have followed a different Markov model or may not considered the entire urban pattern at each location decision. The participants which deviated considerably from the Markov model exhibits partially 'disorganized' patterns with large number of small clusters. The most "none Markovian" participant is rather unique since he produced a distinct urban pattern.

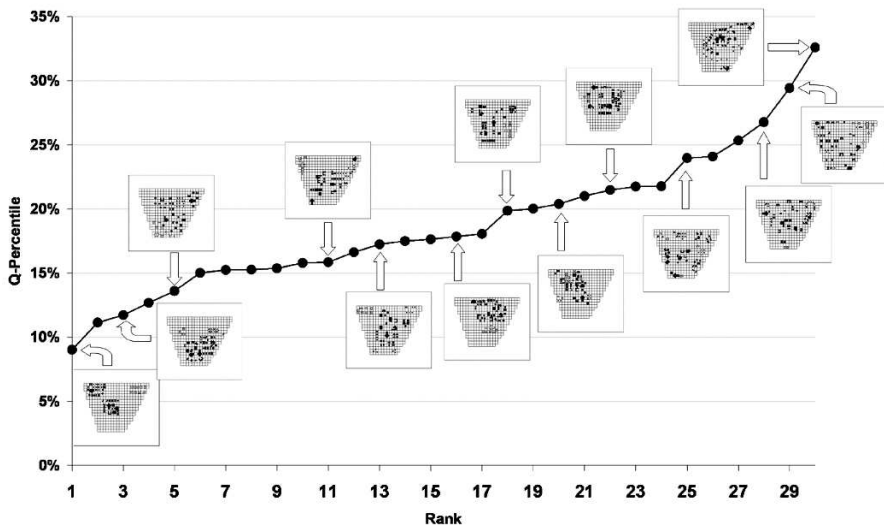


Fig. 9. The values of $\langle q_i \rangle$ for the 30 experiments

Acknowledgements

We thank Hernan Casakin, Department of Architecture, The College of Judea and Samaria for his contributions at the early stages of this research.

References

- Bell EJ (1974) Markov analysis of land use change: an application of stochastic processes to remotely sensed data. *Socio-Economic Planning Science* **8**: 311-316
- Benenson I, Omer I, Hatna E (2002) Entity-based modeling of urban residential dynamics - the case of Yaffo, Tel-Aviv. *Environment and Planning B* **29**: 491-512
- Benenson I, Torrens PM (2004) *Geosimulation: automata-based modeling of urban phenomena*. Wiley, London
- Boerner REJ, DeMers MN, Simpson JW, Artigas FJ, Silva A, Berns LA (1996) Markov models of inertia and dynamism on two continuous Ohio landscapes. *Geographical Analysis* **28**: 1: 56-66
- Brown DG, Pijanowski BC, Duh JD (2002) Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *Journal of Environmental Management* **59**: 247-263
- Clark PJ, Evans FC (1954) Distance to nearest neighbor as a measure spatial relationships in populations. *Ecology* **35**: 445-453
- Gigerenzer G, Goldstein DG (1996) Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review* **103**: 4: 650-669
- Gobin A, Campling P, Feyen J (2002) Logistic modelling to derive agricultural land use determinants: a case study from southeastern Nigeria. *Agriculture, Ecosystems & Environment* **89**: 3: 213-228
- Hathout S (2002) The use of GIS for monitoring and predicting urban growth in East and West St Paul, Winnipeg, Manitoba, Canada. *Journal of Environmental Management* **66**: 229-238
- Isaacson DL, Madsen RW (1976) *Markov chains: theory and applications*. Wiley, New York
- Jahan S (1986) The Determination Of Stability And Similarity Of Markovian Land Use Change Processes:A Theoretical And Empirical Analysis. *Socio-Economic Planning Sciences* **20**: 4: 243-251
- LaGro JA, DeGloria SD (1992) Land use dynamics within an urbanizing non-metropolitan county in New York State (USA). *Landscape Ecology* **7**: 4: 275-289
- Lopez E, Bocco G, Mendoza M, Duhau E (2001) Predicting land-cover and land-use change in the urban fringe A case in Morelia city, Mexico. *Landscape and Urban Planning* **55**: 271-285
- Pentland A, Liu A (1999) Modeling and prediction of human behavior. *Neural Computation* **11**: 229-242
- Portugali J (1996) Inter-representation networks and cognitive maps. In: Portugali J (eds) *The construction of cognitive maps*. Kluwer, Dordrecht, pp 11-43
- Portugali J (2000) *Self-Organization and the City*. Springer, Berlin
- Portugali J (2002) The Seven Basic Propositions of SIRD (Synergetic Inter-Representation Networks). *Nonlinear Phenomena in Complex Systems* **5**: 4: 428-444
- Robinson VB (1978a) Information Theory and Sequences of Land Use: An Application. *Professional Geographer* **30**: 174-179
- Robinson VB (1978b) Land Use Change on the Urban Fringe: A Stochastic Model Analysis. Discussion Paper No 3, Department of Geography, Kent State University, Ohio
- Rubinstein A (1998) *Modeling bounded rationality*. MIT Press, Cambridge, MA

- Seetharaman PB (2003) Probabilistic versus random-utility models of state dependence: an empirical comparison. *International Journal of Research in Marketing* **20**: 1: 87-96
- Weng Q (2002) Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. *Journal of Environmental Management* **64**: 3: 273-284
- Yeh AGO, Li X (2002) A cellular automata model to simulate development density for urban planning. *Environment and Planning B-Planning & Design* **29**: 3: 431-450

Self-Organization and Optimization of Pedestrian and Vehicle Traffic in Urban Environments

Dirk Helbing, Anders Johansson, Stefan Lämmer

Institute for Transport and Economics, Dresden University of Technology, Germany

Abstract

Self-organization is not only a feature of urban evolution, but also found within urban environments. Here, we will focus on three aspects: (i) spatio-temporal patterns in pedestrian flows and their implications for optimized pedestrian facilities, (ii) trail formation of pedestrians and their consequences for improved way systems, and (iii) a self-organization approach to an adaptive control of traffic lights. This chapter will discuss the problems, modelling concepts, results, and solutions, while the mathematical formulation and analysis of the models is presented elsewhere.

1 Spatio-Temporal Patterns in Pedestrian Flows

Dynamical features of pedestrian flows are not only a matter of efficiency and comfort, but also a matter of safety. Since 1945, more than thirty crowd panics have occurred, in which 1,000 people were killed altogether and 3,400 were seriously injured. In 2001 alone, there were two major crowd stampedes in soccer stadia of South Africa and Ghana with over 160 fatalities. Even in the United States, many people died in 2003 due to crowd panics in night clubs on Rhode Island and in Chicago.

1.1 Observed Self-Organization Patterns in Pedestrian Crowds

Problems in pedestrian crowds typically arise due to counterflows, bottlenecks, or intersecting flows. These have been recently studied by video techniques (Daamen and Hoogendoorn 2003). In the following, we will address some of these central issues of pedestrian dynamics in more detail (Helbing et al. 2005a, 2005d).

1.1.1 Counterflows

Under everyday conditions, pedestrians with opposite directions of motion are not equally distributed over cross sections of the walkway. Instead, one observes the separation of pedestrian counterflows into lanes of uniform walking direction (O-

eding 1963; Helbing 1991). This segregation phenomenon reduces the number of encounters with oppositely moving pedestrians, i.e. their interaction frequency and number of necessary braking or avoidance maneuvers. In such a way, the efficiency of walking, i.e. the average velocity into the desired direction of motion is maximized.

In cases of extreme densities, large disturbances, or nervous pedestrian crowds, ordered lanes can break down. Then, blocks of pedestrians with opposite desired directions of motion face each other, but cannot progress. This phenomenon is called “freezing by heating” and may emerge in counterflows of panicking pedestrians (Helbing et al. 2000a).

An unexpected observation is that counterflows can be significantly more efficient than unidirectional flows (Helbing et al. 2005a). Although this is rather surprising, it is consistent with findings by AlGadhi et al. (2002). The reason for this increase in capacity is probably the better coordination between people who meet each other in opposite directions, as they can react to each other (Goffman 1971). In contrast, in unidirectional streams pedestrians do not sufficiently react to what happens behind them (Helbing and Molnár 1995). This causes conflicts (e.g. suppressed overtaking maneuvers) and coordination problems, which reduce the efficiency of motion.

1.1.2 Bottlenecks

In recent experiments, we have investigated mutual obstructions and perturbations at bottlenecks (Helbing et al. 2005a). We found that the relative standard deviation of the time intervals between successive passages of a measurement cross section was more than doubled by the presence of a bottleneck, compared with a similar experiment with no narrowings. The related increase in the level of perturbations reduces the pedestrian speed and flow at the bottleneck. Long bottlenecks seem to be worse than local obstructions.

If opposing flows at a narrow bottleneck disturb each other, we find oscillations in the passing direction. This is reflected by the large variation in the time gaps of successive pedestrians (see Fig. 1). The oscillation effect is more pronounced for longer bottlenecks, which have smaller oscillation frequencies. Typically, it is groups of people rather than single individuals who pass the bottleneck from one side to the other, before people from the other side have a chance to pass the bottleneck. As in the lane formation phenomenon, it is easier to follow someone than to move against an opposing stream. The resulting stream of people releases the pressure in the pedestrian crowd on one side, while the pressure on the other side increases, partially due to impatience. If the pressure difference becomes large enough, the unidirectional stream of people is stopped. Then, people from the other side start to occupy and pass the bottleneck.

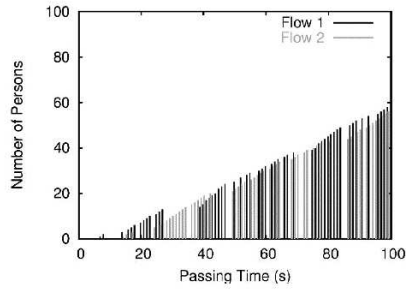


Fig. 1. Passing times and number of passed pedestrians, where opposite flow directions are represented by different shades. The distance between successive bars of the same shade corresponds to the time headway between pedestrians moving into the same direction, while the slope defined by the top ends of the bars determines the pedestrian flow (the number of passing pedestrians per unit time). Compared to the flow through a corridor without narrowings, the pedestrian flow directly after a bottleneck is less regular due to oscillations in the passing direction (see Helbing et al. 2005b). There is a high tendency that the bottleneck is passed by clusters of pedestrians with the same direction of motion rather than by single individuals in an alternating manner.

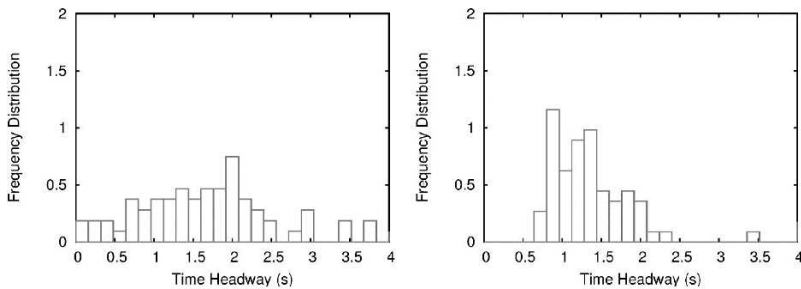


Fig. 2. While the brutto time gap distributions of pedestrians in a bi-directional stream is flat before a long bottleneck (above), it has a pronounced maximum after it (below). This shows that pedestrians keep more equal distances in bottleneck areas compared to less crowded regions, which indicates a kind of condensation or crystallization effect in congested zones (after Helbing et al. 2005a).

In situations with bi-directional flows, bottlenecks should be avoided or, if unavoidable, bottlenecks should be short. Long narrow bottlenecks involve the risk that people try to turn the flow direction in an oscillatory way, but do not manage to do so. In these situations, the flow at the bottleneck drops significantly or stops completely. As time goes by, pedestrians reduce the distance to their predecessors, which eventually leads to a compression of the waiting crowd. This psychologically induced queuing behavior can be frequently observed and mathematically described (Helbing 1991, 1997). The compression in the crowd is related with shock waves (Virkler and Elayadath 1994), which produce the impression that the crowd is moving forward, even if nobody has passed the bottleneck. Therefore,

the people in the back continue trying to get ahead. As the impatience in the crowd increases, people start to push and to interact physically. Consequently, higher and higher pressures build up in the crowd. In extreme situations, this can cause dangerous clogging phenomena due to mutual obstructions, at least temporarily.

1.1.3 Intersecting Flows

Intersections of pedestrian streams are a big problem and practically unavoidable, as it is uncommon to use bridges for their separation. Let us first discuss what happens when two wide streams intersect. In this situation, the phenomenon of stripe formation has been observed (Ando et al. 1988). Like lanes, stripes are a segregation phenomenon, but not a stationary one. Instead, the stripes are density waves moving into the direction of the sum of the directional vectors of both intersecting flows (Dzubiella and Löwen 2002). Naturally, the stripes extend sideways into the direction which is perpendicular to their direction of motion. Therefore, the pedestrians move forward with the stripes and sideways within the stripes. This allows the pedestrian streams to penetrate each other in a way that pedestrians do not have to stop. Lane formation corresponds to the particular case of stripe formation where both directions are exactly opposite. In this case, there is no intersection of flows and motion takes place within more or less stationary stripes (“lanes”).

Previous investigations of intersecting flows (Helbing 1997) indicate that no stable pattern exists when three or more pedestrian streams intersect. Instead, one observes very short-lived patterns which destroy each other, producing a rather chaotic appearance. These temporary patterns include

- rotary traffic in clock-wise or counter-clockwise direction,
- dominant pedestrian flows in opposite directions with the perpendicular directions waiting, and
- short-lived stripes in one of the four possible diagonal directions.

The steady competition and mutual destruction of these spatio-temporal patterns of motion leads to very inefficient pedestrian streams, including temporary blockages.

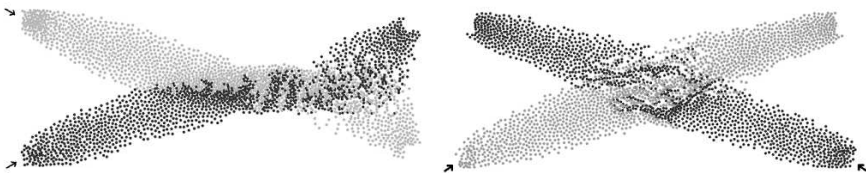


Fig. 3. Stripe formation in two intersecting pedestrian streams. The simulation results were obtained with the social force model (after Helbing et al. 2005a).

1.2 Use of Computer Simulations

The various self-organization phenomena discussed in Sec. 1.1 are a good test for pedestrian simulation models. Any serious evacuation simulation software should be able to reproduce these phenomena, otherwise it will not make reliable predictions. One plausible model, which realistically reproduces these phenomena is the social force model (Helbing 1991, 1997; Helbing and Molnár 1995; Molnár 1996; Helbing and Vicsek 1999; Helbing et al. 2000a, 2000b, 2001). It has been complemented by a route choice module by Hoogendoorn et al. (2002). As lane formation (Helbing 1991, 1997; Helbing and Molnár 1995; Molnár 1996; Helbing and Vicsek 1999; Helbing et al. 2000a, 2001), oscillations of the passing direction at bottlenecks (Helbing and Molnár 1995; Molnár 1996; Helbing 1997; Helbing et al. 2001), as well as clogging and herding phenomena in panicking pedestrian crowds (Helbing et al. 2000b) have been already demonstrated in previous publications, we will restrict ourselves to showing simulation results of stripe formation in intersecting flows, here (see Fig. 3). The model specification has been chosen as usual (see, for example, Helbing et al. 2000b).

The social force model describes the behavior of pedestrians by a superposition of force terms reflecting their different motivations (Helbing 1991, 1997; Helbing and Molnár 1995; Molnár 1996; Helbing and Vicsek 1999; Helbing et al. 2000a, 2000b, 2001). Their wish to move with a certain desired velocity towards a specific destination is delineated by a driving term. Repulsive forces reflect the tendency to keep a certain distance to other pedestrians, obstacles, borders, and dangers. The effect of a stage or of windows displays is described by attractive forces. The same applies to the tendency of group or family members to stay together. These assumptions have been formulated in a mathematical way (Helbing 1991, 1997; Helbing and Molnár 1995; Molnár 1996; Helbing et al. 2000b). Once the model parameters have been calibrated with empirical data of pedestrian streams, the corresponding computer simulations yield, according to our experiences in the past, realistic results even for new geometries and situations, without having to adapt the model or parameters again. That is, the social force model has predictive value, and it allows one to investigate new scenarios, for which experiments would be costly, difficult, or dangerous. This is particularly important for the prognosis, planning, and optimization of evacuation scenarios.

The social force model is used for pedestrian simulations in many countries of the world and has been applied to many practical problems in the meantime. Figure 4, for example, illustrates the evacuation of two decks of a ship or two levels of a hotel. The underlying study has shown that the model is also suitable to treat different personalities (the different characteristic behaviors in emergency situations), complex geometries, and three-dimensional interactions at staircases (Werner and Helbing 2003). Hoogendoorn et al. (2002) have simulated multi-destination flows at Schiphol airport in the Netherlands. With today's computer power, it is even possible to simulate the evacuation of more than 10,000 people (Quinn et al. 2003). The model has also been used to simulate the situation on the Jamarat bridge, which has to cope each year with several million pilgrims within a few days. Figure 5 illustrates the scenario.

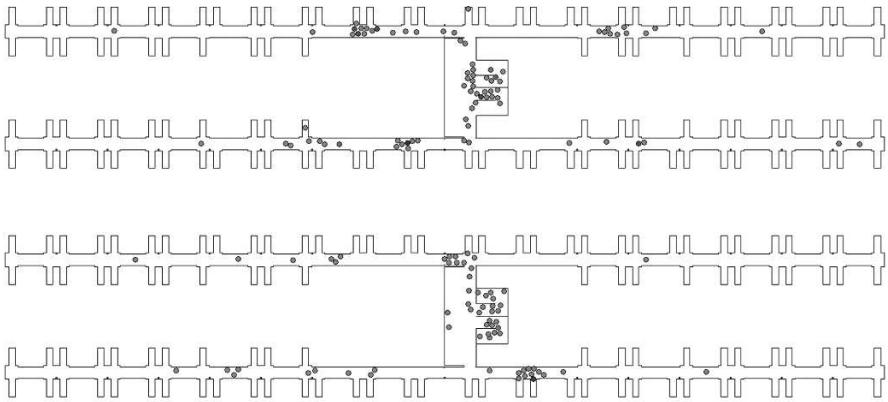


Fig. 4. Snapshot of the computer-simulated evacuation of two floors of a hotel (or two decks of a ship). The corresponding study has demonstrated that the social force model can cope with complex geometries, three-dimensional interactions at staircases, different personalities, and large numbers of pedestrians. For more details of the simulation see Werner and Helbing (2003).

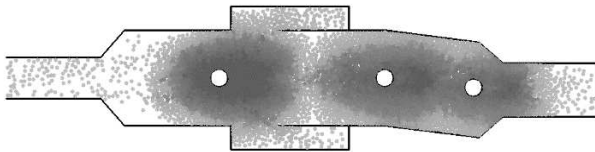


Fig. 5. Simulation of the Jamarat bridge which has to cope with millions of pilgrims each year within a few days (after Helbing et al. 2005a). The pilgrims need to get close to the circular area in order to throw stones at the Jamarah (the pillar in the center). This leads to high pedestrian densities behind the pillars.

The implementation of our pedestrian microsimulator has been recently optimized for efficiency and scalability in terms of computational speed and memory. In the meantime, it is no problem to simulate 30,000 pedestrians in real-time on a normal PC. Parallel computing will even allow one to simulate more than 100,000 pedestrians in real-time. This is suitable for the simulation of large scenarios such as pedestrian flows in extended urban areas (see Fig. 6) or during mass events such as street parades or carnivals (Quinn et al. 2003).



Fig. 6. Densities of pedestrian streams in a simulation of the city center of Dresden, Germany. Our simulation includes 30.000 pedestrians (after Helbing et al. 2005d).

In order to be suitable for scenarios of this complexity, we have implemented certain additional features complementing the social force model:

1. a powerful scenario generation module based on XML,
2. path finding and route choice modules to describe the selection of destinations, including spontaneous stops and impulse shopping,
3. calibration modules to adapt simulations to real measurements,
4. a statistics module evaluating various performance measures (such as densities, velocities, flow efficiency, pressure, overall evacuation time, etc.) to assess and compare different scenarios,
5. an optimization module based on evolutionary algorithms which allows to improve the geometrical boundary conditions of pedestrian facilities,
6. various visualization modes for the illustration of the simulation results, both on- and off-line,
7. a functionality reflecting density-dependent herding effects.

1.3 Some Improved Design Solutions for Pedestrian Facilities

The complex interaction of several pedestrian streams can lead to completely unexpected results, which is a consequence of their non-linear dynamics. Therefore, the planning of pedestrian facilities with conventional methods does not guarantee the avoidance of jams, obstructions, or blockages, particularly not in emergency situations. In contrast, a skillful flow optimization utilizing the self-organization phenomena emerging in pedestrian streams can increase the efficiency and safety of pedestrian facilities at comparable costs and even with less available space.

In the following, we will discuss improved design solutions for intersections and egress routes of stadia, which may also be transferred to other highly frequented places such as hotels, convention centers, exhibition halls, business centers, shopping malls, railway stations, or airport terminals. These are a direct consequence of the various observed self-organization phenomena described above. For more examples of design solutions see Helbing et al. (2005a).

1.3.1 Counterflows

Counterflows are most efficient, when they are organized in a few wide lanes with stable interfaces. At mass events, opposite flow directions must be artificially separated, as the lanes become sensitive to perturbations (Helbing et al. 2000a) and tend to obstruct each other due to attempted overtaking maneuvers. If the lanes break down, dangerous blockages are likely to occur, which may finally trigger crowd stampedes. Surprisingly, the desired separation of opposite flow directions can be supported by a series of obstacles such as railings, trees, or columns (see Fig. 7). Such kinds of solutions are partially permeable, but have psychologically and physically similar effects as a separating wall. They prevent pedestrians from using small gaps in the opposite flow for overtaking maneuvers, which could cause broader streams in one direction than the other, later on provoking an interaction-induced reduction in the walking speed and serious obstructions. However, when the other side is little frequented, a permeable design allows pedestrians to use the other side as well. For this reason, a series of obstacles is flexible and can reduce the frequency of deceleration and avoidance maneuvers. In this way, the average velocity is increased, although some space is used up by the series of obstacles.



Fig. 7. The interfaces of opposite pedestrian flows can be efficiently stabilized by a series of columns (here in a tunnel of the metro system of Budapest at Deak Tér). Nevertheless, the permeability allows for a flexible usage in accordance with the respective pedestrian traffic volumes in both directions (after Helbing et al. 2005a).

Similar solutions are useful at doors: It is often better to have two separate doors for the two opposite directions of motion than one twice as big door. At high pedestrian volumes, each door is typically used by one flow direction only (Mol-

nár 1996; Helbing 1997). In contrast, one door tends to produce oscillatory changes in the passing direction with intermediate periods of standstill during which both directions struggle with each other. This struggling can produce a dangerous pressure in the crowd, which does not build up when there is a reasonable and continuous flow in both directions through separate doors.

1.3.2 Bottlenecks and Clogging in Pushy Crowds

One possibility to improve the situation at bottlenecks is a funnel-shaped design (which, by the way, requires less walkable space). However, a funnel-shaped design is not sufficient in big pedestrian crowds, when large pressure may build up. This is well-known from mass events and may turn them into deadly stampedes. Appropriately placed columns (pillars) can improve the situation. Their functioning is similar to conventional wave breakers. They can absorb pressure in the crowd and reduce it to a subcritical level. However, they do not seriously obstruct the movement towards the exits, as wave breakers (i.e. transversal railings) would do. The performance of various alternative design solutions for exit areas has recently been compared by means of computer simulations (Escobar and de la Rosa 2003). As another example of application, Fig. 8 illustrates an improved design for egress routes of stadia (Helbing et al 2005a).

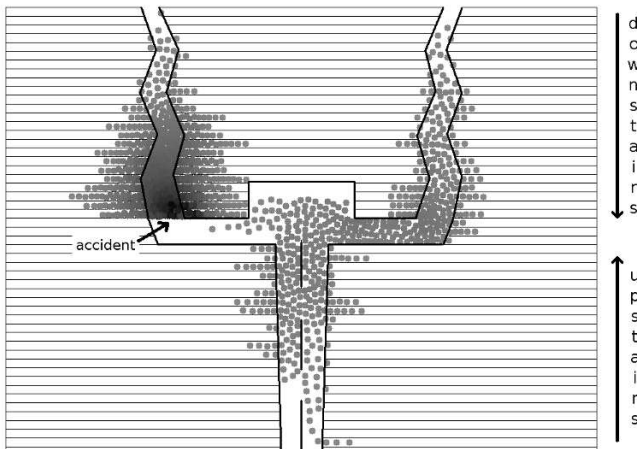


Fig. 8. Improved design of the staircases and an exit of a stadium sector: In the improved design, an increasing diameter of corridors reduces waiting times and impatience (even with the same number of seats), thereby accelerating evacuation. Moreover, the zigzag design of the downwards staircases interrupts the pushing direction in the crowd. This limits the pushing-related build-up of pressure in a crowd and avoids that many fallen people are piled up one on top of each other (after Helbing et al. 2005a).

1.3.3 Intersections

In intersecting flows, mutual obstructions are practically unavoidable. They are, therefore, the greatest challenge to architects, urban planners and organizers of mass events (e.g. Olympic games). In Fig. 3, we have seen that two intersecting flows may form efficient, moving stripes which allow pedestrian streams to penetrate each other without major obstructions. However, when three or more flow directions intersect, the situation is more difficult. In these situations, coherent stripes cannot form, as stripes in several directions would have to propagate through each other, which would destroy any existing stripes. Patterns in intersection areas, therefore, tend to be unstable, “chaotic”, and inefficient.

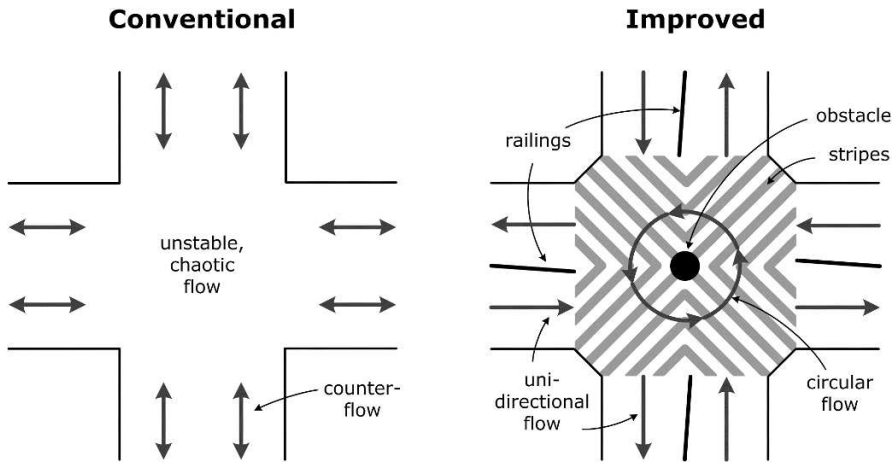


Fig. 9. Left: At conventional intersections, one observes continuously changing patterns of motion, i.e. the pedestrian streams are unstable, “chaotic”, and inefficient. Right: If four intersecting pedestrian flows are separated by barriers, one-way regulations, and an obstacle in the center of the intersection, one can expect the formation of stripes in all four intersection areas of two pedestrian streams, which propagate perpendicularly to their sideward extension (after Helbing et al. 2005a, 2005d). As a consequence, an efficient round-about traffic emerges at high densities. The design makes use of the preference of the right-hand side in Central European pedestrian streams. In other countries such as Japan, it would be reasonable to invert the flow directions.

Nevertheless, the different flows at an intersection (and turning manoeuvres) can be organized in a way that supports a stable pattern. It is remarkable that, in computer simulations, an obstacle in the middle of an intersection has been found to improve the efficiency of motion, although it reduces the available space for the pedestrians (Molnár 1996). This becomes understandable, when the different flow directions are separated in a way that there are four intersections of two flow directions each (see Fig. 9). All four intersection areas are then expected to produce efficient stripes as described above. Good solutions will give rise to rotary traffic,

when the crossing is heavily frequented. In periods of light traffic, some pedestrians may try to move opposite to the main stream. During these periods, however, a deviation from rotary motion does not matter a lot. Our first empirical investigations confirm this assessment.

2 Trail Formation

Another interesting self-organization phenomenon in pedestrian crowds is the development of trail systems on deformable ground. A theory of human trail formation in green areas like public parks must be able to answer the following questions: What are the typical topological structures of trail systems? How and by which mechanism do trail systems evolve in space and time? Why do trails reappear at the same places, if they were destroyed? How should urban planners design public way systems so that walkers actually use them? Why do pedestrians sometimes build trails in order to save three to five meters, but in other cases accept detours which are much larger? Obviously, it is the relative, not the absolute detour that counts. Up to 25% detour seems to be acceptable to pedestrians. If this limit is exceeded, they tend to leave existing ways. However, whether a trail forms or not also depends on the frequency of usage.

Early empirical investigations of human trail formation have, for example, been carried out by Klaus Humpert (Humpert et al. 1996) and Eda Schaur (1991). It is known that many way systems and even streets have originated from human or animal trails. For reasons of easier orientation, these mostly point to optically significant points. However, the pedestrians' tendency to take the shortest way to these points and the specific properties of the terrain are often insufficient to explain the trail characteristics. Although trails serve as shortcuts, they frequently do not supply the shortest way to the respective destination. Rather they are compromises between ways which point into different directions (see Fig. 10).

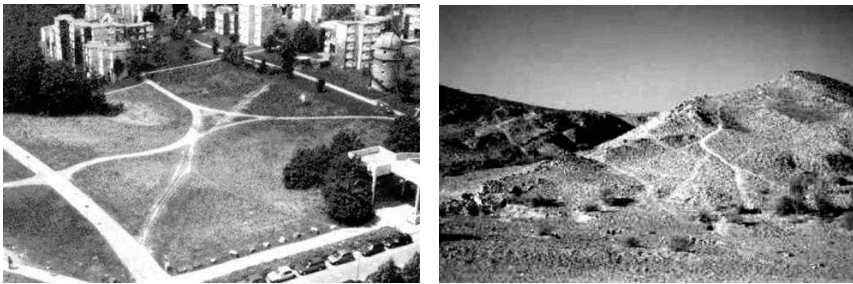


Fig. 10. These examples of human trails (above) and animal trails (below) show that trails are compromises (after Helbing 1997, 1998, Helbing et al. 1997a). They often deviate from the direct way where they meet other trails. The splitting of a trail which meets another way (and thereby gives rise to an island in the middle) is a rather frequently observed phenomenon.

For example, one often observes a splitting of trails, just before they meet another way in a more or less perpendicular direction (see Fig. 10). The minimal angle between splitting ways is found to be about 30° . It appears surprising that no direct way system between the entry points and destinations is formed.

2.1 The Active Walker Model

To simulate the typical features of trail systems, the afore mentioned social force model has been extended by additional equations which describe how pedestrians change their environment and how they are affected by it. The resulting model (Keltsch 1996; Helbing et al. 1997a, 1997b, 2001; Helbing 1998) belongs to the class of *active walker models* (Schweitzer 2003). Like random walkers, active walkers are subject to fluctuations and influences of their environment. However, they are additionally able to locally change their environment, e.g. by altering an environmental potential, which in turn influences their further movement and their behavior. In particular, changes produced by some walkers can influence the other walkers. Hence, this non-linear feedback can be interpreted as an indirect interaction between the active walkers via environmental changes, which may lead to the self-organization of large-scale spatial structures.



Fig. 11. When pedestrians leave footprints on the ground, trails will develop, and only parts of the ground are used for walking (in contrast to paved areas). The similarity between the simulation result (left) and the trail system on the university campus of Brasilia (right) is obvious (photograph by Klaus Humpert) (after [Molnár 1996; Helbing 1997, 1998; Helbing et al. 1997a, 1997b]).

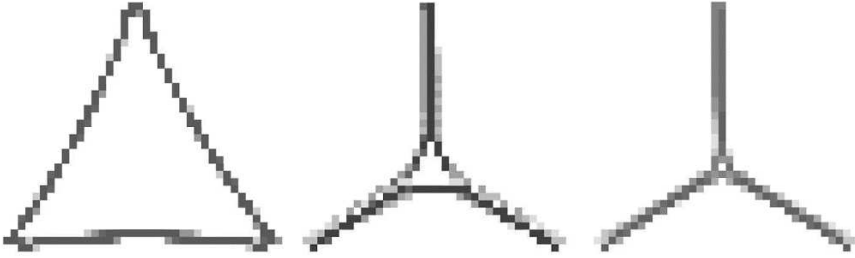


Fig. 12. The structure of the emerging trail system essentially depends on the parameter κ reflecting the attractiveness of existing trails (after Keltsch 1996; Helbing 1997, 1998; Helbing et al. 1997a, 1997b). If κ is small, a direct way system develops (left). If κ is large, a minimal way system is formed (right). Otherwise a compromise between both extremes will result (middle), which looks similar to the main trail system in the center of Fig. 10 (left).

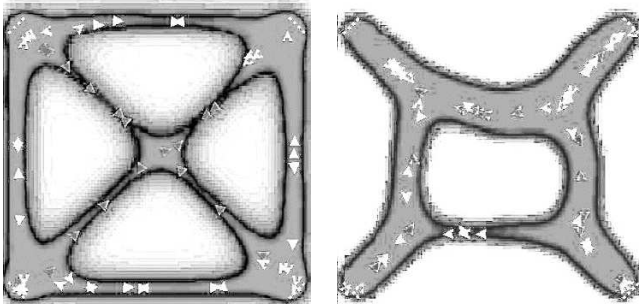


Fig. 13. The left and middle graphics illustrate the trail potential in the beginning and the end of our active walker simulation (after Keltsch 1996; Helbing 1997, 1998; Helbing et al. 1997a, 1997b). Arrows represent the positions and directions of pedestrians. These use, with a frequency ν , each of the six connections between the four entry points and destinations in the corners of a square with edge length L . Starting with a spatially homogeneous ground, the chosen ways change considerable in the course of time. In the beginning, pedestrians take the direct ways (left). Since frequently used trails become more comfortable, a bundling of trails sets in which reduces the overall length of the trail system (middle). The finally resulting way system could serve as a planning guideline (right). It provides a suitable compromise between small construction costs and a large comfort of walking. At an overall length which is 50% shorter than the direct way system, it requires everyone to take a relative detour of 21%, which is a fair and acceptable solution. Moreover, the resulting trail system is structurally stable under the assumption that a frequency 1.5ν of usage is needed to support permanent trails in competition with the vegetation's regeneration. If the frequency of usage is higher than this threshold, a direct way system may be supported.

Active walker models have proved their versatility in a variety of applications, such as formation of complex structures, pattern formation in physico-chemical

systems, aggregation in biological or urban systems, and generation of directed motion (Schweitzer 2003). The approach provides a quite stable and fast numerical algorithm for simulating processes involving large density gradients, and it is also applicable in cases where only small particle numbers govern structure formation. In particular, the active walker model is applicable to processes of pattern formation which are intrinsically dependent on the history of their creation.

Apart from an equation of motion, the active walker model of trail formation contains an additional equation describing how footprints on deformable ground eventually generate trails, thereby increasing the comfort of walking. However, if not regularly used, they eventually fade away due to the regeneration of vegetation. Another equation describes the attractiveness of existing trails as a function of their distance and their visibility. The trail attraction field finally enters an equation for the desired direction of motion, which is assumed to be a compromise between pointing to the destination directly and pointing into the direction of the steepest descent of the gradient of trail attraction.

Our simulations begin with a spatially homogeneous ground. With a certain rate, pedestrians move between given entry points and destinations, starting at a random point in time. In Fig. 11, the entry points and destinations are distributed over the small ends of the ground. While in Fig. 12 (Fig. 13) pedestrians move between all possible pairs of three (four) fixed places.

In the beginning, pedestrians take the direct ways to their respective destinations, because there is no reason to choose another route (see Fig. 13, left). However, after some time they begin to use already existing paths, since this is more comfortable than to clear new ways. By this, a kind of selection process (Schweitzer and Schimansky-Geier 1994) between trails sets in: On the one hand, frequently used trails are more comfortable and, therefore, more attractive than others. For this reason they are chosen very often. The resulting reinforcement makes them even more attractive, until a saturation effect becomes effective. On the other hand, rarely used trails are destroyed by the vegetation's regeneration. This limits the maximum length of the way system which can be supported by a certain rate of trail usage. As a consequence, the trails begin to bundle, especially where different paths meet or intersect. Finally, pedestrians with different destinations use and support common parts of the trail system, which explains the empirically found deviations from direct way systems (see Figs. 12 and 13, middle).

A direct way system (which provides the shortest connections, but covers a lot of space) only develops in cases of high frequencies of usage and almost equal walking comfort for all routes (Fig. 12, left). If the advantage κ of using existing trails is large (like for a rough and rapidly regenerating ground), the final trail system is a minimal way system (which is the shortest way system that connects all entry points and destinations) (Fig. 12, right). For realistic values of κ , the evolution of the trail system stops before this state is reached (Fig. 12, middle). Thus, κ is related to the average relative detour of the walkers.

Note that recent experiments have shown that the predictions of the active walker model of trail formation are in better agreement with observations than Steiner

graphs (Goldstone et al. 2005). The thorough study by Goldstone et al. has just proposed minor improvements of the simulation model.

2.2 Optimization of Way Systems

The trail systems resulting from the above described model are particularly suited as planning guidelines for the construction of optimal way systems: First of all, they take into account the walking and orientation habits, so that pedestrians will actually use such ways. Second, the resulting trail systems provide the best compromise between maximum shortness and comfort of ways. Third, they seem to offer fair solutions, which balance the relative detours of all pedestrians (Fig. 13, right).

Computer simulations of this kind can be used for answering various questions, given a knowledge of the entry points and destinations as well as the expected rates of usage of the corresponding connections (which can be determined by established models (Hoogendoorn et al. 2002)): Which is the trail system that pedestrians would naturally use? Is the resulting way system structurally stable with respect to small changes in the usage pattern? Given a certain amount of money to build a way system of a certain length, which way system should be built, i.e. which one is most comfortable or 'intelligent'? If a certain level of comfort shall be provided, which is the cheapest way system fulfilling this demand? Given an existing way system, how should it be extended?

Due to the large number of walkers and the different time scales involved, the above proposed active walker simulations of trail formation are relatively time consuming. Therefore, equivalent macroscopic equations for trail formation have been derived, which can be solved by a simple and fast iteration scheme. A detailed discussion of this issue has been presented in Helbing et al. (1997b).

3 Self-Organized Traffic Light Control

The optimal online control of traffic intersections is an old, but still not fully solved problem. When the arrival flows of vehicles are low, it is known that the first-in-first-out principle or the right-before-left principle without any further traffic regulations work well. For higher flows, rotary traffic has shown to be efficient, reaching only small delays in travel times. However, for high traffic flows, it is better to bundle cars with the same or compatible directions and to serve them groupwise rather than one by one. This implies an oscillatory mode of service, since it saves clearance times to serve many cars with the same direction. In other words, traffic light control can increase intersection capacity at high arrival flows, but it results in excess waiting times, when traffic volumes are low.

A large amount of effort has been spent on optimizing traffic light control in the past. The classical approaches require vast amounts of data collection and processing as well as huge processing power. Centralized control concepts, there-

fore, imply a tendency of overwhelming the control center with information, which cannot be fully be exploited online. Furthermore, today's control systems have difficult times responding to exceptional events, accidents, temporary building sites or other changes in the road network, failures of information channels, control procedures, or computing centers, natural or industrial disasters, catastrophes, or terrorist attacks.

These weaknesses could be overcome by a decentralized, adaptive approach, which can utilize more local information. Its independence from a central traffic control center promises a greater robustness with respect to localized perturbations or failures and a greater degree of flexibility with respect to the local situation and requirements (Helbing et al. 2005b). In a pending patent, we have described an autonomous adaptive control based on a traffic-responsive self-organization of traffic lights, which leads to reasonable operations, including synchronization patterns such as green waves. In particular, our principle of self-control is suited for irregular (i.e. non-Manhattan type) road networks with counterflows, with main roads (arterials) and side roads, with varying inflows, and with changing turning or assignment fractions. This distinguishes our approach from simplified scenarios investigated elsewhere (Brockfeld et al. 2001; Fouladvand and Nematollahi 2001; Huang and Huang 2003).

3.1 Problems of Optimal Signal Control

Traffic light control shares many features with on-line production control. Both systems can be treated as dynamical queuing networks, in which topological network properties play an important role (Helbing et al. 2004). In order to reduce delays and to increase the throughput of the system, it is necessary to coordinate and synchronize services (e.g. green signal periods) in a suitable way.

Deriving the system optimal control, however, is numerically very demanding. In fact, traffic control is a so-called NP-hard problem: The required computational time to determine the exact optimal solution explodes, i.e. it grows more than polynomially with the number of network nodes. Therefore, even supercomputers fail to solve this optimization problem online for big cities, and the best solution algorithms are based on approximations or heuristic solutions (see the literature discussed in Helbing et al. (2005b)). However, no matter how sophisticated these approaches are,

- either their responsiveness is limited and they appear as tools both coercive and normative (imposing a traffic situation rather than responding to it),
- or they are completely demand-responsive (CLAIRE or PRODYN for instance) and lack a global coordination,

although the strategy proposed at the Technical University of Crete (Diakaki et al. 2003) might be viewed as a reasonable compromise.

A decentralized control approach reduces the computational complexity and the sensitivity to the far remote traffic situation a lot. As there is usually only one globally optimal solution, but a large number of nearly optimal solutions, the idea is

to find the one which fits the local situation and demand best. Nevertheless, coordination between the neighboring areas of decentralized control and even optimization of signal control within local areas remain a scientific challenge. Here, we mention just two of the main problems originating from contradictory requirements:

1. On the one hand, one would like to have long green-time periods in order to reduce efficiency losses due to switching (yellow time periods with no full service). On the other hand, one would like to react immediately to the dynamics at neighboring traffic lights in order to establish synchronization patterns, e.g. a green wave.
2. On the one hand, one would like to use all available space in the road network when the traffic demand is high. On the other hand, the efficiency of queuing systems tends to go down when buffers are getting full, in particular as vehicles may block space required for the service of more important traffic streams. This problem of “parasitic flows” may require to reserve space for the prioritization of certain flows despite of the existing lack of space.

Apart from this, coordination among neighboring areas may profit from delays in service. This paradoxical “slower-is-faster effect” has also been observed in other systems, from pedestrians (Helbing et al. 2000b) to production (Helbing 2004), and recently in intersecting pedestrian-vehicle flows (Jiang et al. 2005).

3.2 Self-Induced Oscillations

Altogether, the optimization of traffic flows in networks and of varying production processes is an extremely difficult subject, which is even hard to imagine in all its aspects and reminds of solving the Gordic node. Our first approach to the problem was inspired by pedestrian flows, see Sec. 1.1.2: In pedestrian counterflows at bottlenecks, one can often observe oscillatory changes of the passing direction, as if the pedestrian flows were controlled by a traffic light. Therefore, we extended this principle to the self-organized control of intersecting vehicle flows, see <http://www.trafficforum.org/trafficlights>.

Oscillations are a organization pattern of conflicting flows which allows to optimize the overall throughput under certain conditions (Helbing et al. 2005c) (see Sec. 3). In pedestrian flows (see Fig. 14), the mechanism behind the self-induced oscillations is as follows: Pressure builds up on that side of the bottleneck where more and more pedestrians have to wait, while it is reduced on the side where pedestrians can move ahead and pass the bottleneck. If the pressure on one side exceeds the pressure on the other side by a certain amount, the passing direction is changed.

Transferring this self-organization principle to urban vehicle traffic, we define red and green phases in a way that considers “pressures” on a traffic light by road sections waiting to be served and “counter-pressures” by subsequent road sections, when these are full and green times cannot be effectively used (Helbing et al.

2005b). Generally speaking, these pressures depend on delay times, queue lengths, or potentially other quantities as well.

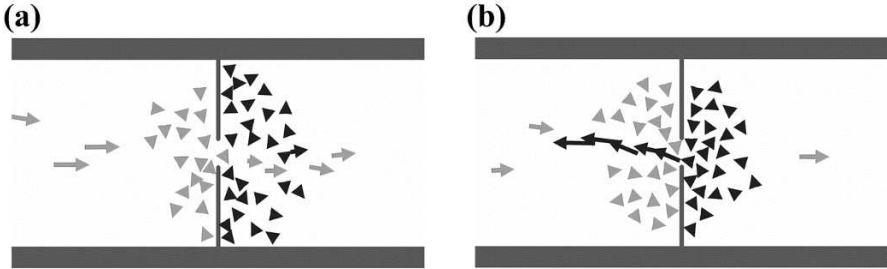


Fig. 14. Alternating pedestrian flows at a bottleneck. These oscillations are self-organized and occur due to a pressure difference between the waiting crowd on one side and the crowd on the other side passing the bottleneck (after Helbing and Molnár 1995; Molnár 1996; Helbing 1997).

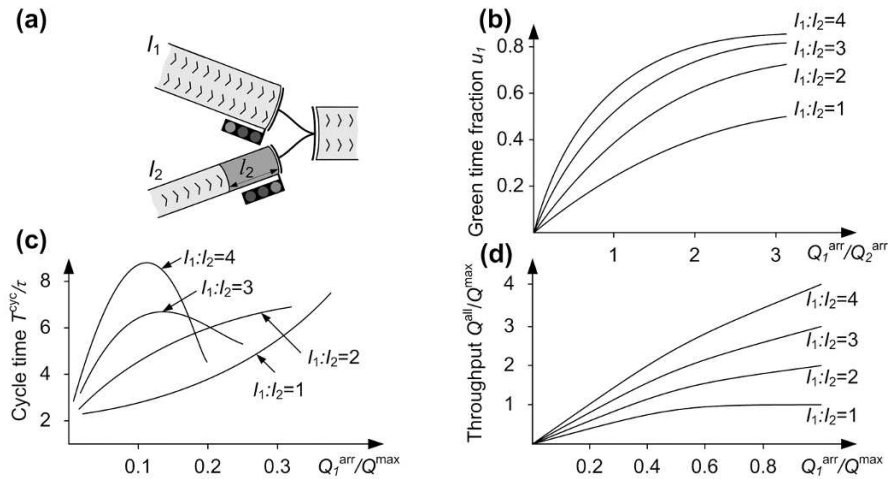


Fig. 15. (a) Illustration of the traffic control of a merge bottleneck for constant arrival rates and a non-congested outflow. The characteristic behavior of the proposed self-organized traffic light control depends on the number I_i of lanes of the entering road sections i and on the arrival rates Q_i^{arr} : (b) Actual green time fraction u_i for $Q_2^{arr} = \text{const.}$ and variable Q_1^{arr} , (c) cycle time T^{cyc} as compared to the yellow time period τ for $Q_2^{arr} = Q_1^{arr}$, and (d) actual throughput Q^{all} of the signalized intersection in comparison with the maximum uninterrupted flow Q^{max} per lane for $Q_2^{arr} = Q_1^{arr}$ (after Helbing et al. 2005b).

The proposed control principle is self-organized, autonomous, and adaptive to the respective local traffic situation. It provides reasonable control results (see Fig. 15), but does not supply a good solution to the coordination problem among neighboring nodes. How to solve this difficult problem is described in our patent.

3.3 Properties of the Self-Organized Signal Control

Our proposed autonomous, decentralized control strategy for traffic flows has certain interesting features (Helbing et al. 2005b): Single arriving vehicles always get a green light. When the intersection is busy, vehicles are clustered, resulting in an oscillatory and efficient service (even of intersecting main flows). If possible, vehicles are kept going in order to avoid capacity losses produced by stopped vehicles. This principle bundles flows, thereby generating main flows (arterials) and subordinate flows (side roads and residential areas). If a road section cannot be used due to a building site or an accident, traffic flexibly re-organizes itself. The same applies to different demand patterns in cases of mass events, evacuation scenarios, etc. Finally, a local dysfunction of sensors or control elements can be handled and does not affect the overall system. A large-scale harmonization of traffic lights is reached by a feedback between neighboring traffic lights based on the vehicle flows themselves, which can synchronize traffic signals and organize green waves. In summary, the system is self-organized based on local information, local interactions, and local processing, i.e. decentralized control.

We should point out some interesting differences compared to conventional traffic control:

- The green phases of a traffic light depend on the respective traffic situation on the previous *and* the subsequent road sections. They are basically determined by actual and expected queue lengths and delay times. If no more vehicles need to be served or one of the subsequent road sections is full, green times for one direction will be terminated in favor of green times for other directions. The default setting corresponds to red lights, as this enables one to respond quickly to approaching traffic. Therefore, during light traffic conditions, single vehicles can trigger a green light upon arrival at the traffic signal.
- Our approach does not use precalculated or predetermined signal plans. It is rather based on self-organized red and green phases. In particular, there is no fixed cycle time or a given order of green phases. Some roads may be even served more frequently than others. For example, at very low traffic volumes it can make sense to serve the same road again before all other road sections have been served. In other words, traffic optimization is not just a matter of green times and their permutation.
- Instead of a traffic control center, we suggest a distributed, local control in favor of greater flexibility and robustness. The required information can be gathered by optical or infrared sensors, which will be cheaply available in the future. Complementary information can be obtained by a coupling with simulation models. Apart from the section-based traffic model (Helbing 2003; Helbing et

al. 2005b) favored by us, one can also use other (e.g. microsimulation) models with or without stochasticity, as our control approach does not depend on the traffic model. Travel time information to enhance route choice decisions may be transmitted by mobile communication.

- Pedestrians could be detected by modern sensors as well and handled as additional traffic streams. Alternatively, they may get green times during compatible green phases for vehicles or after the maximum cycle time T^{\max} . Public transport (e.g. busses or trams) may be treated as vehicles with a higher weight. A natural choice for the weight would be the average number of passengers. This would tend to prioritize public transport in a natural way. In fact, a prioritization of public transport harmonizes much better with our self-organized traffic control concept than with precalculated signal plans.

4 Summary and Conclusions

We have discussed many empirical observations in pedestrian crowds pointing to a variety of self-organization phenomena. These include lane-formation, oscillations of counterflows at bottlenecks, formation of stripes in intersecting flows, and clogging phenomena of pushing crowds at bottlenecks. Self-organization phenomena can be used to improve the design of pedestrian facilities, e.g. of intersections or egress routes of stadia.

Moreover, we have sketched the features of a universal pedestrian simulator which is presently being developed. It allows one to simulate evacuation scenarios, but also pedestrian flows in extended urban areas. This simulator is based on the social force model, a wide-spread microscopic pedestrian model. Applications to the optimization of pedestrian facilities are quite natural. A clever use of the self-organized patterns of motion even allows one to reach more efficient pedestrian flows with less space. In addition, optimized way systems can be generated with an active walker model of human trail formation, an extension of the social force model. This model includes additional indirect interactions between pedestrians based on the consideration of their environmental impact and its feedback on human walking behavior.

Finally, we have discussed the problem of optimal signal control of urban traffic networks. The model used by us (Helbing 2003; Helbing et al. 2005b) allows one to efficiently simulate the transitions between free and congested traffic, taking into account congestion-responsive traffic assignment and adaptive traffic control. The proposed decentralized control concept is ideally suited to utilize new sensor and wireless communication technologies, which are becoming cheaply available. Our adaptive signal control is inspired by self-organized oscillations at bottlenecks found in pedestrian flows. It is flexible, robust, and decentralized rather than based on precalculated signal plans and a vulnerable traffic control center. It is expected that this new control philosophy will increase the quality of service and also enter the realm of on-line production scheduling and coordination

of organization processes, as it is more suited to satisfy local demands at a high system performance than classical, centralized control concepts.

Acknowledgments

This work has been partially supported by the German Research Foundation (DFG projects He 2789/5-1 and He 2789/7-1). S.L. thanks for his scholarship by the “Studienstiftung des Deutschen Volkes”. The authors are also grateful for inspiring discussions and fruitful cooperation with Lubos Buzna, Illés Farkas, Tilo Grigat, Joachim Hahne, Joachim Keltsch, Christian Kühnert, Péter Molnár, Torsten Werner, and Tamás Vicsek.

References

- AlGadhi SAH, Mahmassani HS, Herman R (2002) A speed-concentration relation for bi-directional crowd moments with strong interaction In: Schreckenberg M, Sharma SD (eds) *Pedestrian and Evacuation Dynamics*. Springer, Berlin, pp 3-20
- Ando K, Oto H, Aoki T (1988) Forecasting the flow of people. *Railway Research Review* **45**: 8-14
- Batty M, Desyllas J, Duxbury E (2003) The discrete dynamics of small-scale spatial events: agent-based models of mobility in carnivals and street parades. *Int. J. Geographical Information Science* **17**: 7: 673-697
- Brockfeld E, Barlovic R, Schadschneider A, Schreckenberg M (2001) Optimizing traffic lights in a cellular automaton model for city traffic. *Physical Review E* **64** (056132)
- Daamen W, Hoogendoorn SP (2003) Experimental Research of pedestrian walking behaviour. *Transportation Research Records* **1828**: 20-30
- Diakaki C, Dinopoulou V, Aboudolas K, Papageorgiou M, Ben-Shabat E, Seider E, Leibov A (2003) Extension and new applications of the traffic-responsive urban control strategy: Coordinated signal control for urban networks. *Transportation Research Board* **1856**: 202-211
- Dzubiella J, Löwen H (2002) Pattern formation in driven colloidal-mixtures: tilted driving forces and re-entrant crystal freezing. *J. Phys.: Cond. Mat.* **14**: 9383-9396
- Escobar R, de la Rosa A (2003) Architectural Design for the Survival Optimization of Panicking Fleeing Victims. Paper presented at the 7th European Conference on Artificial Life, Dortmund, Germany, September 14–17, <http://www.dd.chalmers.se/armando/panic/report.htm>
- Fouladvand ME, Nematollahi M (2001) Optimization of green-times at an isolated urban crossroads. *European Physical Journal B* **22**: 3: 395-401
- Goffman E (1971) *Relations in Public: Microstudies in the Public Order*. Basic, New York
- Goldstone RL, Jones A, Roberts MA (2005) Group path formation. Preprint Indiana University <http://cognitn.psych.indiana.edu/rgoldsto/pdfs/paths.pdf>
- Helbing D (1991) A mathematical model for the behavior of pedestrians. *Behavioral Science* **36**: 298-310
- Helbing D (1997) *Verkehrsdynamik*. Springer, Berlin

- Helbing D (1998) Pedestrian Dynamics and Trail Formation In: Schreckenberg M, Wolf DE (eds) *Traffic and Granular Flow '97*. Springer, Singapore, pp 21-36
- Helbing D (2003) A section-based queueing-theoretical model for congestion and travel time analysis in networks. *Journal of Physics A* **36**: L593-L598
- Helbing D (2004) Traffic dynamics. In: Radons G, Neugebauer R (eds) *Nonlinear Dynamics of Production Systems*. Wiley-VCH, Weinheim, pp 85-105
- Helbing D, Molnár P (1995) Social force model for pedestrian dynamics. *Phys. Rev. E* **51**: 4282-4286
- Helbing D, Vicsek T (1999) Optimal Self-Organization. *New Journal of Physics* **1**: 13: 1-17
- Helbing D, Keltsch J, Molnár P (1997a) Modelling the evolution of trail systems. *Nature* **388**: 47-49
- Helbing D, Schweitzer F, Keltsch J, Molnár P (1997b) Activer walker model for the formation of human and animal trail systems. *Physical Review E* **56**: 2527-2539
- Helbing D, Farkas I, Vicsek T (2000a) Freezing by Heating in a Driven Mesoscopic System. *Physical Review Letters* **84**: 1240-1243
- Helbing D, Farkas I, Vicsek T (2000b) Simulating dynamical features of escape panic. *Nature* **407**: 487-490
- Helbing D, Molnár P, Farkas I, Bolay K (2001) Self-organizing pedestrian movement. *Environment and Planning B* **28**: 361-383
- Helbing D, Witt U, Lämmer S, Brenner T (2004) Network-induced oscillatory behavior in material flow networks and irregular business cycles. *Physical Review E* **70** (056118)
- Helbing D, Buzna L, Johansson A, Werner T (2005a) Self-organized pedestrian crowd dynamics: experiments, simulations and design solutions. *Transportation Science* **39**: 1-24
- Helbing D, Lämmer S, Lebacque JP (2005b) Self-organized control of irregular perturbed network traffic. In: Deissenberg C, Hartl RF (eds) *Optimal Control and Dynamic Games*. Springer, Berlin
- Helbing D, Schönhof M, Stark HU, Holyst JA (2005c) How individuals learn to take turns: emergence of alternating cooperation in a congestion game and the Prisoner's Dilemma. *Advances in Complex Systems* **8**: 1: 87-116
- Helbing D, Johansson A, Buzna L (2005d) Pedestrian dynamics and evacuation: Empirical results and design solutions. In: Mahmassani HS (ed) *Flow, Dynamics and Human Interaction*. Proceedings of the 16th International Symposium of Transportation and Traffic Theory (ISTTT 16), Elsevier Science, Amsterdam
- Hoogendoorn SP, Bovy PHL, Daamen W. (2001) Microscopic pedestrian wayfinding and dynamics modeling. In: Schreckenberg M, Sharma SD (eds) *Pedestrian and Evacuation Dynamics*. Springer, Berlin, pp 123-154
- Huang DW, Huang WN (2003) Traffic signal synchronization. *Physical Review E* **67** (056124)
- Jiang R, Helbing D, Shukla PK, Wu QS (2005) Inefficient emergent oscillations in intersecting driven many-particle flows. Preprint <http://arXiv.org/abs/cond-mat/0501595>.
- Humpert K, Becker S, Brenner K, Gangler A (1996) Entwicklung grossstädtischer Agglomerationen: Stadtfeldtypen und Stadtfeldcollagen. In: Teichmann K, Wilke J (eds) *Prozeß und Form "Natürlicher Konstruktionen"*. Der Sonderforschungsbereich 230. Ernst & Sohn, Berlin, pp 172-193
- Keltsch J (1996) Selbstorganisation von Wegen durch "Active Walkers". Master's thesis, Universität Stuttgart

- Molnár P (1996) Modellierung und Simulation der Dynamik von Fußgängerströmen. Shaker, Aachen
- Oeding D (1963) Anlagen des Fußgängerverkehrs, Straßenbau und Straßenverkehrstechnik. Bundesministerium für Verkehr, Abt. Straßenbau, Bonn
- Quinn MJ, Metoyer RA, Hunter-Zaworski K (2003) Parallel Implementation of the Social Forces Model. In: Galea ER (ed) Pedestrian and Evacuation Dynamics. Proceedings of the 2nd International Congress CMS Press, London, pp 63-72
- Schaur E (1991) Ungeplante Siedlungen. Krämer, Stuttgart
- Schweitzer F (2003) Brownian Agents and Active Particles. Springer, Berlin
- Schweitzer F, Schimansky-Geier L (1994) Clustering of “active” walkers in a two component system. *Physica A* **206**: 3-4: 359-379
- Virkler MR, Elayadath S (1994) Pedestrian Density Characteristics and Shockwaves. In: Akçelik R (ed) Proceedings of the Second International Symposium on Highway Capacity, Vol. 2. Transportation Research Board, Washington, D.C., pp 671-684
- Werner T, Helbing D (2003) The Social Force Pedestrian Model Applied to Real Life Scenarios. In: Galea ER (ed) Pedestrian and Evacuation Dynamics 2003. Proceedings of the 2nd International Congress. CMS Press, London, pp 17-26

Multidimensional Events in Multilevel Systems

Jeffrey Johnson

Department of Design and Innovation, The Open University, Milton Keynes, UK

Abstract

Design involves assembling parts into whole to satisfy specified relationships. Binary relations between pairs of elements are not rich enough to represent the generality of this. Powerful though they are, graphs and networks must be extended to *multidimensional networks* to represent the many subtleties of complex systems. These structures underlie a natural theory of multilevel systems that has within it a kind of structural time defined by multidimensional *events*. Slow changing relational structure is discriminated from fast moving events captured by patterns of numbers. It will show that the former acts as a kind of multilevel multidimensional *backcloth* for the multilevel multidimensional dynamic *traffic* of the latter. The theory will be illustrated by examples.

1 Introduction

From many perspectives, towns and cities are complex multilevel systems. They have structure between the micro-level of streets and buildings, the meso-level of neighbourhoods, and the macrolevel of the city and its hinterland. It is difficult to say precisely where the boundary of a city lies. My house lies on the boundary between the city of Milton Keynes and the neighbouring county of Bedford. Even though the boundary is well defined as the centre of the road (Fig. 1a), it is artificial at the level of the street. When crossing from one side to the other there is no obvious feeling of moving from the 'urban' system of Milton Keynes to the 'rural' system of mid-Bedfordshire. It seems paradoxical that something as concrete as a city can have such indeterminate functional boundaries. But lack of clear boundaries is a characteristic of many systems. The artist Escher made a speciality of exploring such ambiguities as one system merges into another (Fig. 1b).

How can we plan cities if we don't know where they begin and end? This paper is concerned with ways of making systems less ambiguous, both in terms of their boundaries, their interactions with external systems, and in terms of their internal dynamics. This is an issue of *representation*. As we develop our representation for large multilevel systems, it will become clear that it has within it an implicit structural notion of time which underlies a kind of structural dynamics. These structures define multidimensional *events* in which structural change corresponds to the swing of the pendulum as it measures out clock time (Atkin 1981).



Fig. 1. Ambiguous system boundaries: a) The boundary of Milton Keynes marked by the white line in the road, b) Escher's ambiguous boundary between sky and water¹

2 Relations and Structure: Graphs and Networks Are Not Rich Enough

A *graph* is a set of dots called *nodes*, with some joined together by lines called *links*. The nodes are also called *vertices* and the links are also called *edges*. Nodes and links give a very effective graphical way of representing objects and relationships between them. For example in Fig. 2a the binary relation of mutual love links Romeo and Juliet. Figure 2b shows a *directed* relationship which establishes an asymmetric link between the houses, with the owner of one house buying another house.

In the UK it is common for house sales to be linked in a *chain*, where the sale of House-3 depends on the sale of House-2 which depends on the sale of House-1. Sometimes these chains can quite long, and sometimes the sale of a very expensive house at the end of the chain can be held up by problems with the sale of a very inexpensive house at the start of the chain. It is not unknown for the person owner of the house at the top end of the chain to buy the house at the bottom.

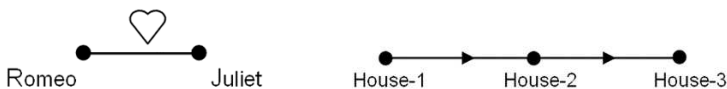


Fig. 2. Links and arrows can represent relationships

Chains of relational structure often underlie unexpected dependencies in systems. For example, it is now known that poisons like mercury can accumulate in the food chain: phytoplankton and bacteria → insects and zooplankton → small fish and aquatic animals → predatory fish → birds and mammals → man. One of

¹ M.C. Escher's 'Sky and Water I', © 2004 The M.C. Escher Company - Baarn - Holland. All rights reserved.

the fundamental ideas in this paper is that *relational structure can support the transmission of system phenomena along chains of connection.*

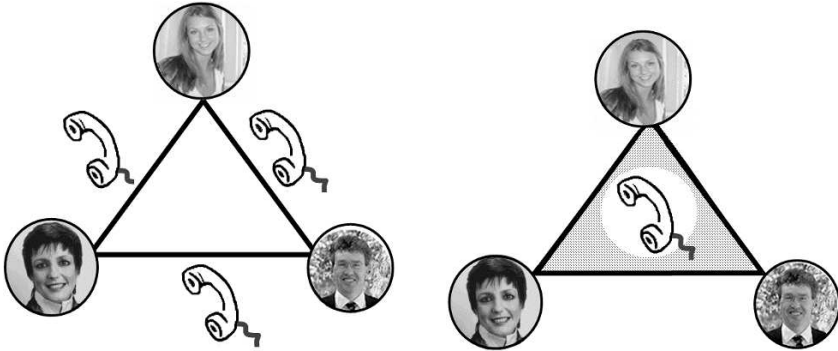


Fig. 3. Binary relations are not rich enough to represent all relational structure: a) three 2-ary relations, b) a 3-ary relation

However, it is essential to understand that binary relations are a special case of a more general kind of relationship that binds *sets* of things together, not just pairs. This is illustrated in Fig. 3, which shows three members of a family, the mother, father, and their young daughter. In Fig. 3a the three people can communicate by phone in pairs, supporting exchanges like “Can I visit my friend? – I’ll ask your mother”, “can she stay with her friend? – where does her friend live – I didn’t ask – you’re hopeless, I’ll call her myself”, “which friend is it and where do they live?”, “it’s a new friend that I met last night and he has a really cool flat in town” – “I think I need to discuss this with your father”, “....”, “OK, how will she get there”, “Can you drive her there”, “OK”, “*!@*! Her phone is engaged”, ... “Hi, I’ll be there to pick you up in half an hour”. Here the pair-wise link structure is not rich enough to allow unobstructed communication.

In contrast if all three people had a conference call as in Fig. 3b, the communication could be much shorter with less chance of misunderstanding: “Can I visit my friend?” - “which friend is it and where do they live?” - “it’s a new friend that I met last night and he has a really cool flat in town” - “....”, “OK, how will you get there”, “I’ll be there to pick you up in half an hour”.

Figure 4 shows a group of seven people having dinner together. Such dinners are characterised by lots of conversations between pairs of people, which sometimes engage another people, forming a 3-ary and 4-ary relations, and sometimes everyone joins in under a 7-ary relation. In the Fig. 4a snapshot there are two binary conversations and one 3-ary conversation. The dynamics and variety of these interactions are usually enjoyable, which is why we like dinner parties, instead of restricting our interactions to just one other person, or dining alone. Some people are even more gregarious, as illustrated by the 400,000 that interacted together at Woodstock in 1969 (Fig. 4b).



Fig. 4. Structure formed by n -ary relations, $n > 2$, are very common in human systems. a) A dinner in which n -ary relations form dynamically though time, $n = 1, 2, \dots, 7$, b) The 1969 Woodstock festival involved a 400,000-ary relation

2 Simplicial Complexes and Emergence

Figure 5 illustrates the generality of representing n -ary relations graphically. If a line can represent a binary relation, a triangle can represent a 3-ary relation, a tetrahedron can represent a 4-ary relation, and an n -hedron can represent an n -ary relation. A line is a 1-dimensional polyhedron, a triangle is a 2-dimensional polyhedron, a tetrahedron is a 4-dimensional polyhedron, and in general an n -ary relation can be represented by an n -hedron in $(n-1)$ -dimensional space.

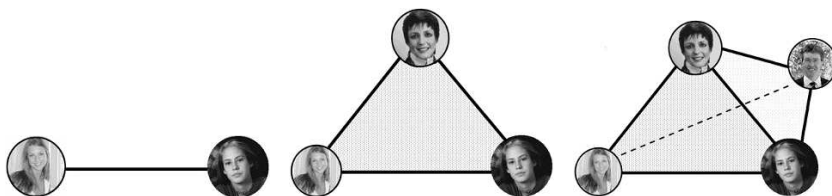


Fig. 5. n -ary relations represented by multidimensional polyhedra in multidimensional space. a) binary relation, b) 3-ary relation and c) 4-ary relation

The polyhedra are also called *simplices* and their elements are called *vertices*. A p -dimensional simplex has $p + 1$ vertices. If $\{v_1, v_2, \dots, v_n\}$ is a set of vertices, we use the notation $\langle v_1, v_2, \dots, v_n \rangle$ to show that they are related under an n -ary relation, and form a simplex. A set of simplices with all their faces is called a *simplicial complex*.

In general, structures formed from n -ary relations have *emergent properties*, that is they have properties that are not possessed by their elements or sub-structures. For example, when coffee, milk, and sugar are mixed they form a drink with a taste which is different from the tastes of the constituent element. Thus $\{\text{coffee, milk, sugar}\} \neq \langle \text{coffee, milk, sugar} \rangle$.

3 Hierarchical Aggregation

The concept of emergence is closely tied up with the idea of multilevel systems. If the vertices of a simplex exist at one level, then any relational simplex involving those vertices exists at the next level up. Thus the mapping from the set to the simplex moves up the hierarchy of representation, from Level N to Level $N+1$, as shown in Fig. 6. In Fig. 6a a face emerges as the features are assembled by the relation, and in Fig. 6b the possibility of passing through the arch emerges from assembling the blocks in this way.

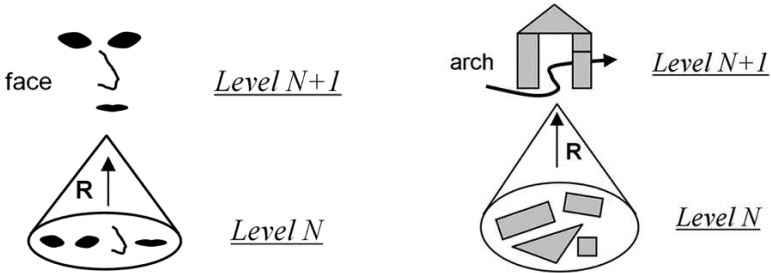


Fig. 6. The hierarchical cone construct of multilevel systems. a) features assembled to form a face, b) blocks assembled to form an arch

Figure 6 illustrates the *hierarchical cone construction*. The *base* of the cone is the set represented by an Euler circle (often shown in perspective as an ellipse), and the *apex* is the structured set (often represented by its name).

The cone construction illustrates a number of interesting and important possibilities. Figure 7a shows that the same set can be assembled in different ways. Thus the set of vertices alone is not sufficient to represent a simplex. For full knowledge we need to know the relation, and therefore we use the notation $\langle v_0, v_1, \dots, v_n; R \rangle$, which provides information on both the vertices of the simplex and on the relation that assembles those vertices into the structure.

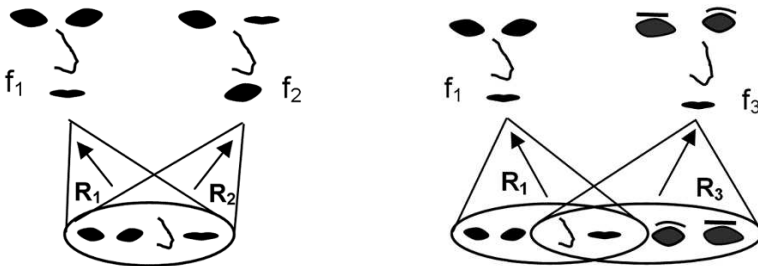


Fig. 7. Hierarchical cones representing assembly of parts into holes. a) relational cones with a shared base, b) relational cones with intersecting cone bases

Figure 7b shows that the bases of hierarchical cones may intersect, and this leads to what we will call a lattice hierarchy in the next section. These intersections are fundamental structures in systems. They are sites of interaction and support the system dynamics. They are the generalisation of shared vertices that connect links in graphs and networks.

4 Alpha- and Beta-Aggregations in Lattices Hierarchies

Whereas there is no doubt that we use n -ary relations to build objects out of their parts, and that this establishes hierarchical levels, there is a subtlety in hierarchical aggregation involving another kind of aggregation. This is illustrated in Fig. 8 where three arches are assembled from their components. These assemblies require *all* the parts for their n -ary relation to hold. We call this an α -aggregation, or an *AND-aggregation*. At the next level the arches are gathered up to form a set. In this case A-1 or A-2 or A-3 is sufficient for an arch. We call this a β -aggregation, or an *OR-aggregation*. Thus the set of arches is defined by a disjunction of conjunctions, $v_j \langle v_j1, \dots, v_jn ; R_j \rangle$.

The hierarchical structure in Fig. 8 can be used to make a diagram in which higher level objects are placed higher on the page than lower level objects, as shown in Fig. 9. A line joins the elements of sets with the structures they aggregate into. Since we allow that lower level objects can be aggregated into more than one higher level object, we do not obtain the classical inverted tree hierarchy. Instead we obtain a *lattice* hierarchy, as shown in Fig. 9.

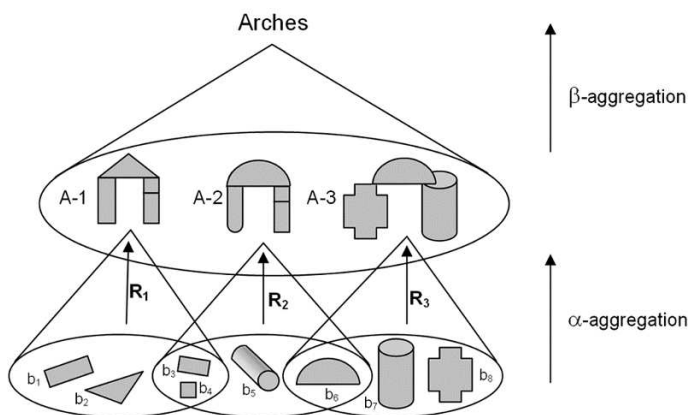


Fig. 8. Two different types of multilevel aggregation

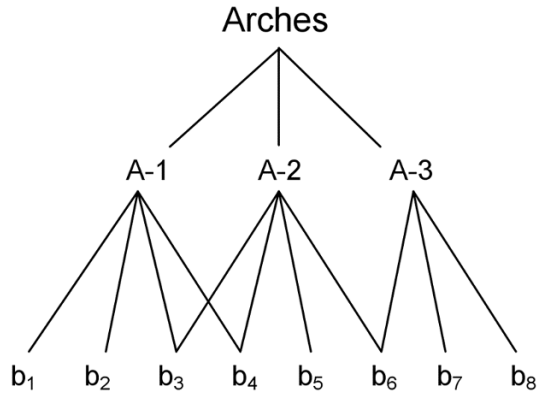


Fig. 9. A lattice hierarchy

5 What Is a City?



Fig. 10. A panorama of part of the city of Paris

What does it mean to say that Paris is a city? In some sense the word Paris is the name of a very complex structure. As we look at the picture of Paris in Fig. 10 we can pick out some of the parts of the city. For example, it has *buildings*, *roads*, *trees*, *the Eiffel Tower*, *monuments*, *parks*, and much else besides. Although they can't be seen in this picture, Paris has a *river*, *traffic*, *a Metro*, *restaurants*, *shops*, *churches*, *bridges*, *railway stations*, and so on. We know that many of the buildings contain *apartments*, that there are *boulangeries* where one can buy *baguettes* and *cakes*, and *cafés* where one can buy *coffee* and *croissants*. The churches contain *paintings* and *sculptures*. And there are over nine million *people*, doing thousands of things.

This illustrates the *intermediate word problem* (Fig. 11). The analysis of any system begins with a notion of 'the system' at some highest level of representation, and a prior collection of all sorts of 'stuff' jumbled up at some lower level. It

is not correct to call this collection a set, because it contains a jumble of ideas and objects at all levels. We call it the *hierarchical soup*, “a prelogical primordial source containing the building blocks of all subsequent structures” (Gould et al. 1984). The intermediate word problem is that of picking out words from the soup, giving them precise meaning, and establishing how the elements and objects they represent are assembled under n -ary relations in a lattice hierarchy.

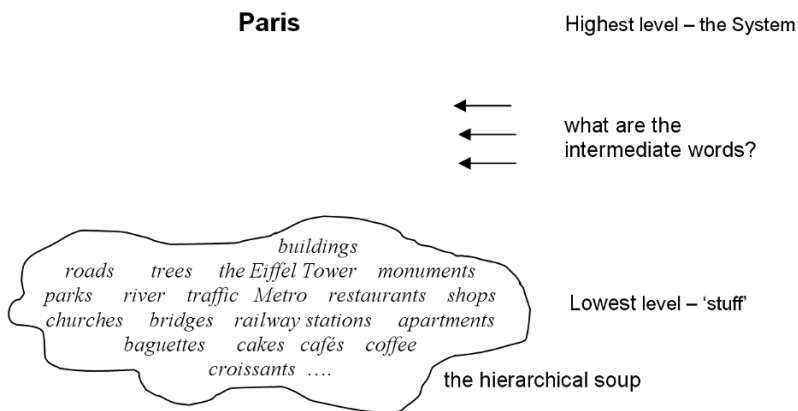


Fig. 11. The intermediate word problem of complex systems

The intermediate word problem is an attempt to create a *symbolic representation* of the city. To do this it is useful to define *stuff* to be anything that can be part of the whole. Thus in Fig. 12a a set of unspecified ‘stuff’ is assembled under the relation R_{Paris} to form the city. In this context the word is acting as an *uninstantiated variable*, much as the symbol x acts as an uninstantiated number in algebra.

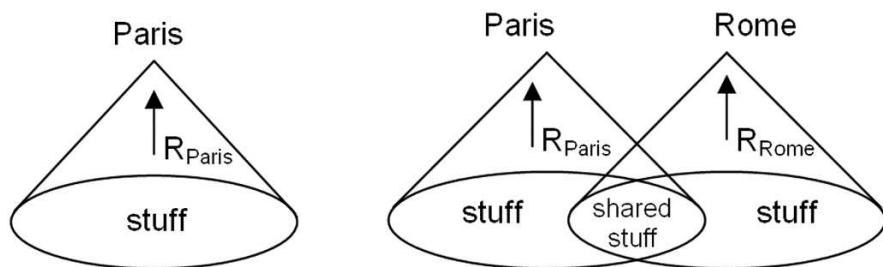


Fig. 12. Seeking intermediate words. a) uninstantiated structure, b) does shared structure define the city?

When trying to understand what makes up a city, one supposes that all cities have some stuff in common. For example, many cities have large rivers running through them, often a factor in their establishment, along with other topographic features. Paris and Rome share the presence of grand classical municipal buildings, which do not occur in my local modern city of Milton Keynes. Thus in

trying to ‘unpack’ the city we can find a set of common vertices in the soup, and begin to bring these together to form polyhedra in the multilevel representation.

The intermediate word problem can work in both top-down and bottom-up ways. Top-down, the city is described by higher level more abstract structures, whose ‘stuff’ is not well instantiated. Bottom-up tends to involve assembling particular things into recognisable units.

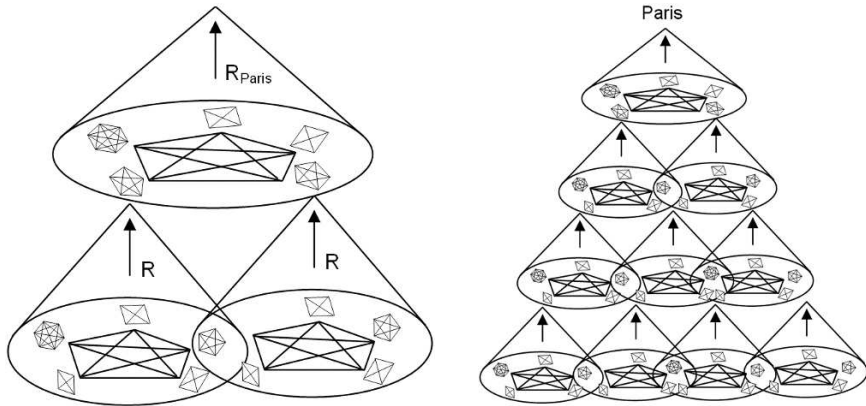


Fig. 13. Intermediate word instantiation of the stuff that makes up cities: a) two levels, b) four levels



Fig. 14. Bottom-up assembly of minor features at street level: a) Belgravia, London, b) Camden, London, c) Tower of London

Figure 14 shows some of the lower level structure of a city, at the level where people live their daily lives. The houses in Belgravia are regular structures with characteristic columned portals and symmetric windows on four levels. In fact by looking at the picture one can conduct an intermediate word analysis, as shown in Fig. 15. Here I have abstract the set {roof, small window, medium window, large window, arched window, double portico}, plus a few more elements such as the linear ledges and the single porticos. Also the visual features are arranged in a way that approximates the buildings in Fig. 14a. This assembly illustrates a very regular structure typical of that style of design.

In contrast to the buildings in Fig. 14a, those in Fig. 14b are much more heterogeneous in form. One of them even has a aeroplane attached to its façade! Even

so, if one looks carefully at these building, they too have a lot of common structure, having also been once part of a terrace in which many windows are the same size and arranged symmetrically in the façade.

The Tower of London shown in Fig. 14c is also has visually interesting structure. The intermediate words include the turrets at the corners rising above the height of the walls that connect them. Here there are windows of various types, crenellated walls and, of course, a flag. Having deconstructed the building into these intermediate words, it could then be reconstructed in a similar way to Fig. 15.

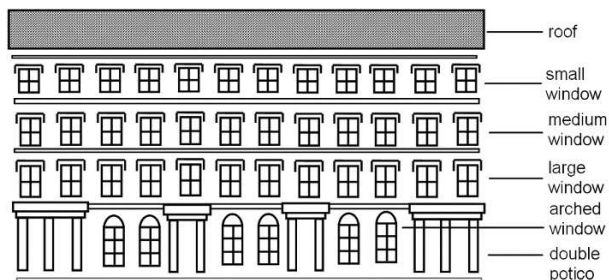


Fig. 15. Abstraction of intermediate words and relational assembly of named features

At the other end of the scale to this local structure, the intermediate words might be generated from maps, in a top-down decomposition of the city into zones. For example, the map in Fig. 16 shows the central part of London. Most people on seeing this map mentally subdivide it into areas of comprehensible size, such as Regent's Park, Covent Garden, Mayfair, Hyde Park and Kensington. These are the level that local planning authorities work at, using combinatorial structures, either implicitly or explicitly.



Fig. 16. A map of London ready for top-down decomposition into zones

One of the great problems in planning is the lack of “joined up thinking”. What the planners do in one zone may not take into account what planners are doing in adjacent zones, and the policies may be incompatible. This is because the multile-

vel representation is not properly developed, and the relationships between the levels in planning is implicit.

6 Multidimensional Connectivity and Q-Analysis

Relational structure is everywhere, simplices are everywhere, and simplices have interesting connectivity properties. Figure 17 shows how simplices can share different numbers of vertices, and that the more vertices they share, the more *highly connected* they are.

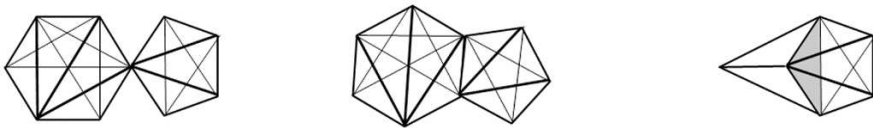


Fig. 17. Simplices can be connected at different dimensions: a) 1 shared vertex, b) 2 shared vertices and c) 3 shared vertices

The intersection of two simplices is called their *shared face*. If the shared face has dimension q , the simplices are said to be q -near. Thus the simplices in Fig. 17a are 0-near (a single vertex has dimension zero), those in Fig. 17b are 1-near (two vertices make a onedimensional line), and those in Fig. 17c are 2-near (three vertices make a twodimensional triangle). We say two simplices are q -connected if there is a chain of pairwise q -near simplices between them.

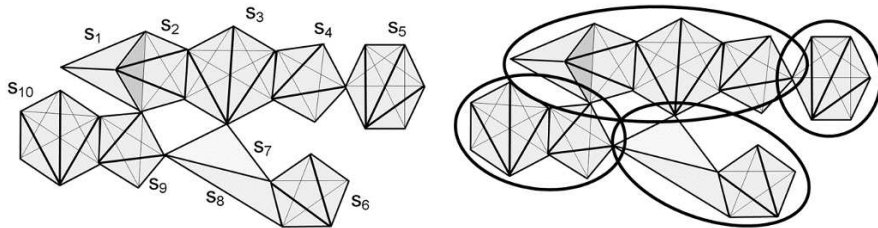


Fig. 18. A simplicial complex of connected simplices: a) the simplicial complex (b) components connected by lines

Table 1 shows what we call the *shared vertex matrix*, which shows the number of shared vertices between the simplices in Fig. 18a. From this we can list the q -connected components as shown in Table 1. For example, there are four distinct components at $q = 1$, which become connected as a single component at $q = 0$ (Fig. 18b).

Figure 19 shows the relationship between four English public houses and the customers that frequent them. Typically people like to go to more that one pub for the variety it brings. Suppose that someone who likes The Swan, the simplex on the left, knows a very good joke. When he gets to the Swan pub he may tell it to

the people who happen to be in that day. They may tell the joke to other people in the pub, and it is likely to be transmitted to everyone in the Swan before the day is finished. The next day, one of those people in the Swan might visit the Anchor, and tell the joke there. Again the joke gets transmitted within the pub. The next day one of the people from the Anchor might visit the Goat pub, and tell the story there. In this way the joke can get transmitted from the Swan pub to the Bull pub, even though they have no customers in common. This illustrates how human information can pass through social structure determined by relations. In general the more highly connected the structure, the more rapidly information is transmitted.

Table 1. Shared vertex matrix

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
s_1	4	3	0	0	0	0	0	0	1	0
s_2	3	4	2	0	0	0	0	0	1	0
s_3	0	2	5	1	0	0	1	0	0	0
s_4	0	0	1	4	1	0	0	0	0	0
s_5	0	0	0	1	6	0	0	0	0	0
s_6	0	0	0	0	0	6	2	0	0	0
s_7	0	0	1	0	0	2	3	2	1	0
s_8	0	0	0	0	0	2	2	2	5	2
s_9	1	1	0	0	0	0	0	0	5	2
s_{10}	0	0	0	0	0	0	0	0	2	6

Table 2. Q-analysis

q day	(3 shared vertices)	$\{s_5\}, \{s_{10}\}$
$q = 4$	(5 shared vertices)	$\{s_2\}, \{s_3\}, \{s_4\}, \{s_5\}, \{s_6\}, \{s_5\}, \{s_{10}\}$
$q = 3$	(4 shared vertices)	$\{s_1\}, \{s_2\}, \{s_3\}, \{s_4\}, \{s_5\}, \{s_6\}, \{s_5\}, \{s_{10}\}$
$q = 2$	(3 shared vertices)	$\{s_1, s_2\}, \{s_3\}, \{s_4\}, \{s_5\}, \{s_6\}, \{s_5\}, \{s_{10}\}$
$q = 1$	(2 shared vertices)	$\{s_1, s_2, s_3, s_4\}, \{s_5\}, \{s_6, s_7, s_8\}, \{s_5\}, \{s_{10}\}$
$q = 0$	(1 shared vertices)	$\{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_5, s_{10}\}$

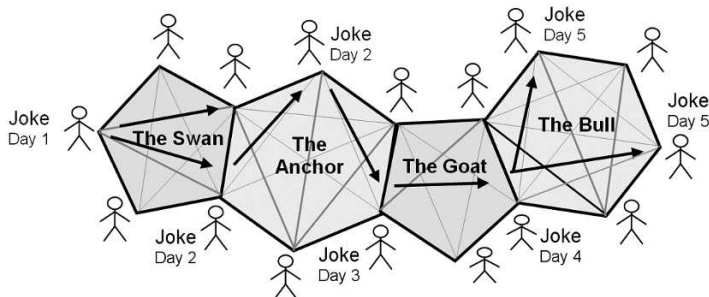


Fig. 19. The transmission of jokes and information on the pub-customer structure

7 Backcloth and Traffic in Multilevel Systems

In the previous section we saw how information can be transmitted across social structure. In cases like this the social structure is relatively constant, while the information traffic is relatively dynamic. Thus the social structure acts as a kind of relatively static *backcloth* that supports a relatively dynamic *traffic* of system activity.

Since the backcloth is formed from multilevel relations, the traffic over it is also multilevel. One of the reasons that people find complex systems difficult to understand is that in their models the traffic does not aggregate properly over the backcloth. People refer vaguely to micro-, meso- and macro-levels, but these are never precisely or operationally defined. Within these vague generality various mapping and functions are analysed, but the way micro-level mapping aggregate to macro-level mappings is rarely considered. Thus micro- meso- and macro- levels are not unified within the theory, which consequently has little or any analytic power.

In Fig. 20 the set of services is related by being in the same zone. Quite naturally, all the spending in the zone is the sum of all the spending on the particular services in the zone.

Formally we say that the set and relational structure at any level is part of the backcloth of the system, and any mappings of the structure into numbers is called *traffic*. Generally backcloth is infrastructure that is expensive to create and changes slowly. In contrast the dynamics of the traffic are represented by changes in their values, and these happen over relatively short time periods.

Figure 21 shows a system in which factories share different job skills. Each factory has a ‘wage for the job’, and the wages for people with each skill are aggregated in to the total wage bill. At the level of the company, the managers try to resolve the cost traffic with the earnings traffic, trying to produce profit traffic.

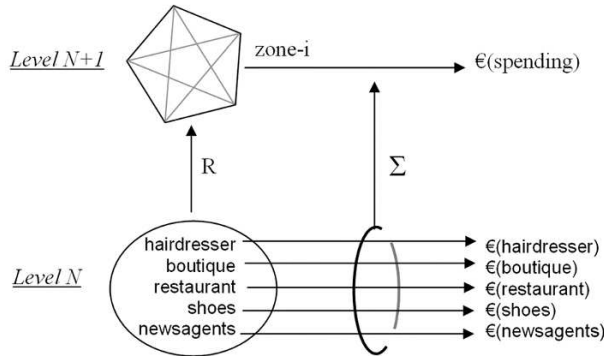


Fig. 20. Hierarchical traffic aggregating over the hierarchical backcloth

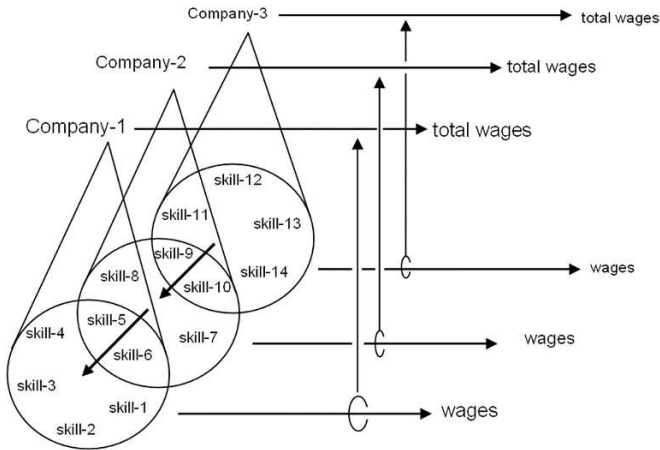


Fig. 21. Order-1 traffic dynamics through the backcloth

Suppose that Company-3 is expanding and requires more worker of all kinds. Then this creates a demand for workers with skill-9 and skill-10 that can pressurise Company 2 to increase wages for these skills. In the short term this may disrupt established differentials in Company-2, leading the other workers to argue for higher wages, including those with skill-5 and skill-6. This in turn puts pressure on Company-1, which may also have to increase its wages. Thus Company-1 may have been induced to change its wages due to company-3, even though they share no job skills and may occupy completely different parts of the market.

Changes in traffic in time are called *Order-1 events*. The peaks and troughs of time series represent order-1 events. Although some understanding may be gained from trying to find patterns in time series data, in most systems the reason for change is due to other kinds of dynamics at lower levels in the system.

8 Polyhedral Events and Order-2 Dynamics

Although we think that dynamics are played out in *clock time* against physical events such as the swing of a pendulum or the daily revolution of the earth, urban systems may have their own time determined by structural system events. Figure 22 shows the event of forming the arch, where a distinction can be made between the arch not having been built, and the arch having been built. We define the formation of such structures as *events*. In other words, every time a simplex is born, it marks an event.

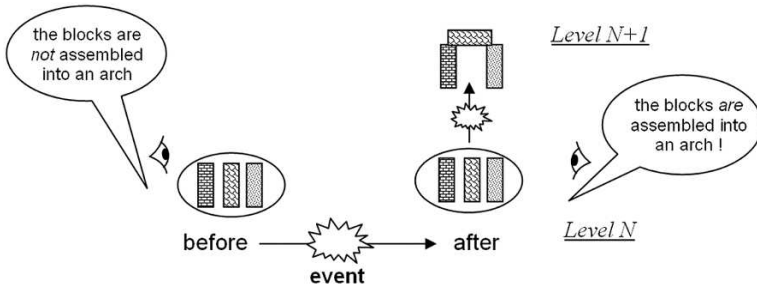


Fig. 22. Changes in structure mark events in system time

Planning involves the identification of key events leading towards an objective, itself usually a composite of many events. For example, a large building project will have many stages, beginning with someone having the idea for the project, collecting data, feasibility planning, raising finance, and beginning to clear the site. These are represented by simplices in Fig. 23, but more likely each event is a complicated multilevel multidimensional structure.

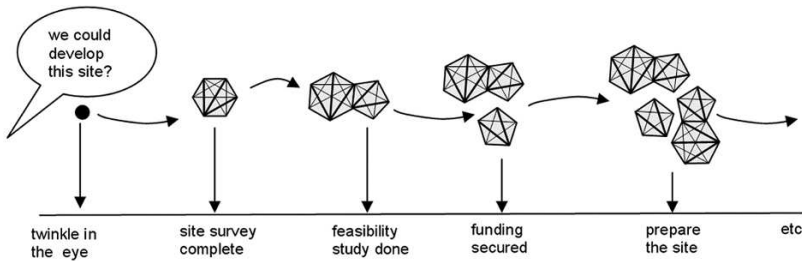


Fig. 23. Planning as a trajectory of events

In an ideal world design, planning and fabrication follow a sequence of well defined activities, with one predictably following another through time. In reality, plans can only attempt to approach this ideal because some events take longer (or shorter) to come to fruition than expected, or because unexpected events intrude into the system and have to be managed. System events tautologically map out system time and dynamics, *e.g.* the apartments can be occupied when after the ‘building is finished’ event, the bridge can carry traffic after the ‘bridge is finished’ event, the wall can be painted after the ‘plaster has dried out’ event.

Which time is most relevant to multilevel dynamics, clock time measured by physical events or system time measured by system events? The answer in most cases is both, and furthermore it is necessary to try to understand how these two times interact. For most systems, it is necessary to wait for one system event, before attempting to create the next. The event of the walls being built usually has to precede the event of the roof being built. However, clock time plays an important part in implementing design because resources are consumed in clock time. People get hungry in clock time and expect to be paid in clock time. Machinery is hired in

clock time, rent is paid in clock time, and interest is paid in clock time. A significant mismatch between the estimated clock time for system events to unfurl and the actual clock time taken can cause projects to fail. Thus, as illustrated in Fig. 24, mapping system events into clock time is an important part of planning and designing complex systems.

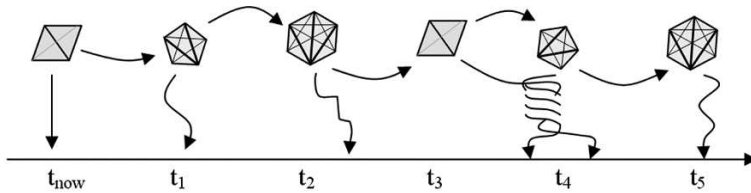


Fig. 24. Mapping system events into clock time

Let a *plan* be defined to be a *trajectory* of system states and events. Then it might be supposed that the plan has to be mapped into clock time as illustrated in Fig. 24. However, in many cases there are many possible futures and the plan can never be finished. For example, London has been planned for centuries, but it will never reach a final state. The plan cannot have a ‘finished’ system state as goal. The plan involves trying to predict, understand, and manage the possible system trajectories.

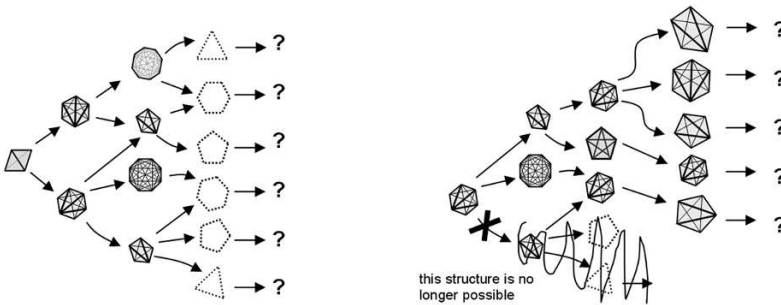


Fig. 25. A trajectory through ever-changing possible words: a) possible future at time t , b) possible future at time $t + \Delta t$

Thus planners have to know what are the possible future system states, how desirable or undesirable those states are, and how to steer the system towards the desirable states, minimising the possibility that the system will be attracted to particularly undesirable states. In Fig. 25 the system starts at its present state, which is rarely where one would want to be². From this position it is possible to envisage future system states, and how the trajectory might evolve, or be steered, towards

² There is a story that a traveller, lost on country roads, asks a country yokel the best way to get to London. The yokel replies that if he was going to London, he wouldn’t start from here.

them. As time moves on some states or transitions become impossible, and large parts of the once-possible future disappear. For example, if a planner decides to allow the cutting down of woods to build houses, any future that involves those woods as an amenity becomes impossible.

Let the *state* of a system be defined to be its instantaneous set of elements, relations, and mappings. When the relational backcloth of a system changes we will call this *Order-2* dynamics, as distinct from changes to traffic which have been defined to be *Order-1* dynamics. As we have seen, planners have to manage multilevel multidimensional trajectories and both the *Order-1* and *Order-2* dynamics associated as the system moves on a trajectory through time.

9 Examples

9.1 Visual Structure in Urban Systems

In Fig. 10 we showed an image of Paris, a city thought by many to be beautiful on many dimensions. Why can't all cities be as beautiful as Paris. Is it simply that this city inherits a wealth of buildings and treasures from the past which others do not. Is it that this city is particularly well designed in terms of its multilevel multidimensional relations?

An intermediate word analysis of Paris soon shows that it is not homogeneously beautiful, and that Paris can compete with other cities worldwide for having some of the ugliest urban environments. As Paris is subdivided into its *arrondissements* a finer perspective becomes possible. For example, Paris-1 is a simplex whose vertices include a wealth of beautiful buildings including the Louvre. All eight of the central *arrondissements* are well endowed with historical buildings, and most of them have river views and bridges. However, some of the outer *arrondissements* are less glamorous, with simplices that have ugly industrial and apartment buildings and derelict plots as vertices.

As a thought experiment, imagine yourself on an open top double-decker tourist bus in Paris with a GPS-enabled portable computer. Suppose that your bus driver shows you all of Paris, and not just the parts on the tourist trail. Every five minutes the bus stops and you look around and give your evaluation of the city as you see it. Your reactions might range from the sublime to the unspeakably bad, and you tap this into the computer which logs your reaction and location. Figure 26 shows what your evaluation map might look like.

In some way, your evaluation of 'Paris' as 'beautiful' or otherwise is an aggregate of these lower level evaluations. How individuals do these evaluations is not known, and this is an important question in environmental science. In my opinion mapping these evaluations to numbers and forming a numerical aggregate would be inappropriate. The cognitive mechanisms are much more complex than that. More likely the overall evaluation is a highly non-linear combination of discrete pieces of information with ordinal weightings and operations that allow some things to be ignored and others to take prominence. In all case, to study such eva-

luations it is necessary to have a theory of multilevel systems that can aggregate evaluations from the micro-micro-level of looking at a flower in the Gardin de Luxembourg, to the macro-macro-level of booking a ticket to what some think is one of the most beautiful cities in the world.

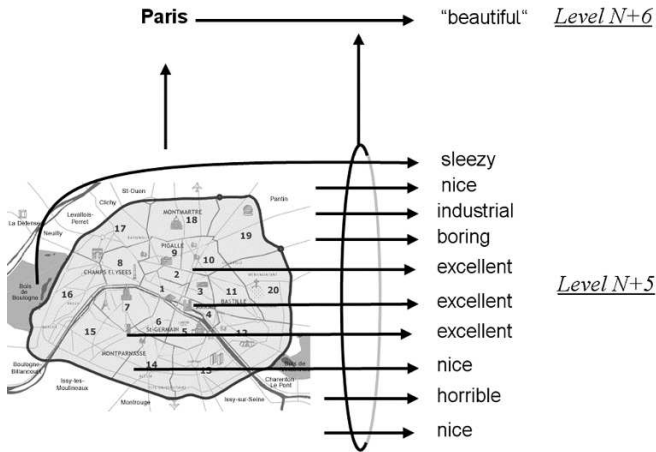


Fig. 26. Evaluating Paris as ‘beautiful’ city as aggregate of lower level evaluations

9.2 Analysis at the Level of Buildings

Figure 27 shows the terrace of four houses in Belgravia discussed in Section 5. Let the intermediate words be the following:

Then each house can be seen as a simplex as follows:

$\sigma(\text{house-1}) = \langle \text{roof, 3 small window, 3 medium window, 3 large window, arched window, double portico} ; R_1 \rangle$

$\sigma(\text{house-2}) = \langle \text{roof, 3 small window, 3 medium window, 3 large window, arched window, single portico} ; R_2 \rangle$

$\sigma(\text{house-3}) = \langle \text{roof, 3 small window, 3 medium window, 3 large window, arched window, single portico} ; R_3 \rangle$

$\sigma(\text{house-4}) = \langle \text{roof, 3 small window, 3 medium window, 3 large window, arched window, double portico} ; R_4 \rangle$



Fig. 27. The terrace seen as a set of houses

A q -analysis of these simplices shows that they are very highly connected, with $\sigma(\text{house-1})$ and $\sigma(\text{house-4})$ being ‘identical’, and $\sigma(\text{house-2})$ and $\sigma(\text{house-3})$ being ‘identical’ in terms of their shared vertices. The relation R_1 is not identical to R_4 , and the relation R_2 is not identical to R_3 .

Although the pairs of houses share the same visual features they are mirror images of each other. Does this matter? This is like asking if it matters that the letter **b** is a mirror image of the letter **d**, or that **d** is the letter **p** rotated by 180 degrees. It depends on the context, and it is relative to the purpose of the system.

Such observations lead to the question what the intersection of two structures actually is. Whereas it is easy to write $\langle a, b, c, d, e, f, g \rangle \cap \langle d, e, f, g, h, i, j \rangle = \langle d, e, f, g \rangle$, what does it mean to write $\langle a, b, c, d, e, f, g; R_1 \rangle \cap \langle d, e, f, g, h, i, j; R_2 \rangle$? A simple solution to could be to write $\langle d, e, f, g; R_1 \cap R_2 \rangle$, leaving open the question of what $R_1 \cap R_2$ might mean.

9.3 Example: A Theory of Design for Urban Planning

Successful design involves taking the right pieces and putting them together the right way. The design process involves discovering sets of the right elements, and relations that will assemble them in desirable ways. At the beginning of the design process there is the idea that one want to build something, but often the precise nature of that something is not clear. For example, in Fig. 28 the designer has a vague idea about redeveloping an area, but the details have yet to be decided.

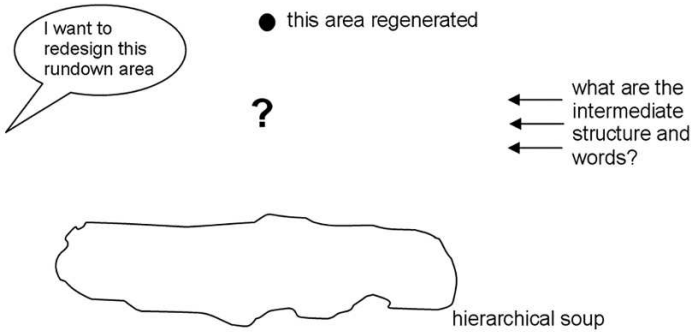


Fig. 28. Design begins with an intermediate word problem

If the designer is an experienced urban planner, his or her hierarchical soup will already be very rich, containing words that represent the things already present, some of which must go. There will be many words representing remembered cityscapes seen during a lifetime, many words representing the technicalities of the planning process, and so on.

Design is both a directed activity and a serendipitous activity. So suppose that Fred’s Diner is established in the development area, and is to be left alone. This part of the design is therefore fully instantiated, and everything else has to work around it. Suppose that the intermediate words include houses and shops. These words are instantiated with particular houses and shops, and they act as uninstantiated variables. And suppose there’s going to be ‘something which is fun’. This intermediate phrase is very vague and could be instantiated with many things.

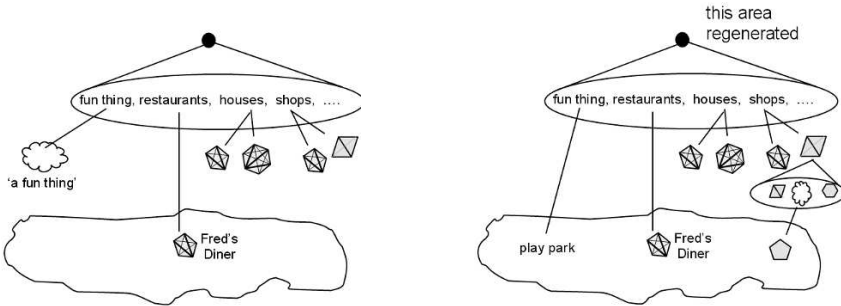


Fig. 29. Design proceeds by filling in intermediate structure: a) filling in detail during the design, b) an abstract concept becomes instantiated

Figure 29 shows two more stages in the design process, with detail being added as the design proceeds. The abstract idea of a ‘fun thing’ is shown as a cloud, waiting to be firmed up. Ultimately the design becomes fully instantiated, when every part of it is *grounded* in the soup. At this stage all the elements of the design are decided, and the ways that they will be assembled as a multilevel multidimensional system have been decided.

Of course design does not work in a linear way, from beginning to end with no changes. As designers try new ideas, an x here or a y there, they implicitly assume that everything will fit together without problems. At the more abstract level there may be no obvious reason to think that x and y might be incompatible. But as more details become instantiated at more levels, tensions may become apparent. Sometimes the numbers don't fit together, e.g. there should be m houses, but if the fun thing is a play park, there won't be room for that many houses – what to do? Sometimes the designer will solve these problems, and sometimes the problems may be unsolvable. When they are unsolvable the designer has to backtrack, to a level in the lattice hierarchy where the conflict is resolved. Then the process of building lower level structure begins all over again until, all the elements and relations are instantiated to become the design blueprint.

9.4 Example: Urban Zones and Road Traffic Systems

When trying to understand towns and cities it is very common to use nested sequences of zones, as illustrated in Fig. 30.

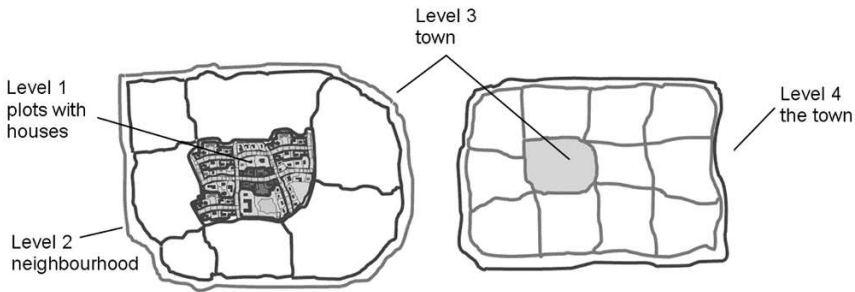


Fig. 30. A hierarchical zoning scheme

Zones like this can be used to develop a multilevel representation of road systems. At the lowest level are the road as links within a zone, as shown in Fig. 32. Then routes between origins and destinations are defined in the usual way. At the next level, Level $N+1$, more abstract links are defined between the boundary nodes of the zones. Each of these links represents all the N -level routes between that pair of nodes. Generally there are many routes to each higher level link, and this simplifies the combinatorially explosion considerably.

This method of moving up the hierarchy works at all levels, with $N+k$ links forming $N+k$ routes, and $N+k+1$ links across zones representing the set of $N+k$ routes between the nodes. The great advantage of the representation is that it is *self-similar*, so that aggregation from Level $N+k$ to level $N+k+1$ is the same as the aggregation of Level $N+k+1$ to Level $N+k+2$.

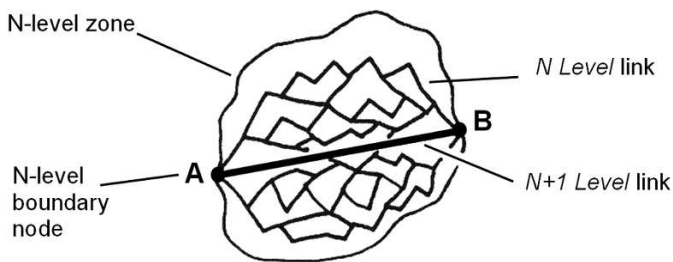


Fig. 31. An (N+1)-Level link is a set of *N*- routes between two boundary nodes

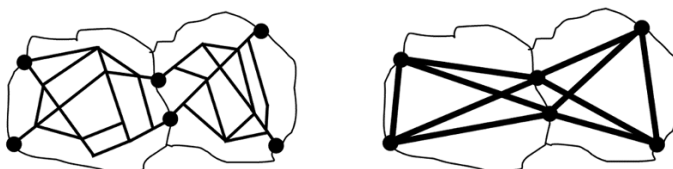


Fig. 32. Hierarchical links defined by hierarchical zones: a) routes made of N-level links, b) N+1 level represent many routes

Figure 33 shows how the representation can be used to represent large numbers of “routes on the ground” by routes made of hierarchical links. This greatly reduces the combinatorial explosion in the number of routes, and gives the possibility of representing road systems of any size, such as the whole of mainland Europe. It can be shown that this representation reduces the computational complexity of calculating routes (Serras 2005).

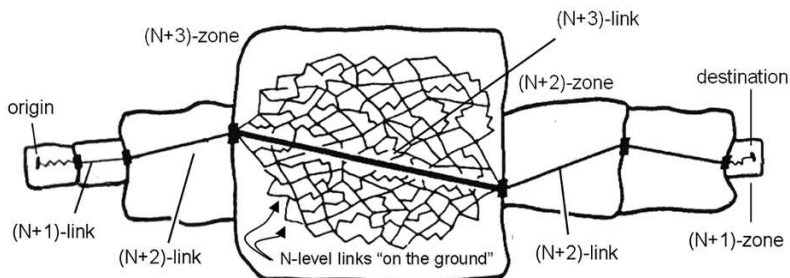


Fig. 33. A hierarchical route made up of hierarchical links

This kind of multilevel multidimensional representation is important in road systems and other networks, such as the internet, because a shortest path is a system-wide phenomenon. A computational complexity of the order of N^2 will preempt treating the system holistically. Dijkstra’s algorithm has computational complexity of the order of $N \log N$ which is much better, but it has disadvantages such

as calculating all shortest paths (even when there is no demand) and requires the whole network and related calculations to be stored as the computation progresses. Can new representation like ours lead to computation complexity less than the order of $N \log N$, say of the order of N or better?³

10 Conclusions

In this paper we have skimmed over many ideas relating multilevel multidimensional systems. The central idea has been that elements can be assembled into structures, and that structures exist at a higher level in the representation than their constituent elements. Furthermore, we have asserted that the formation of structures are systems *events* that mark system time. In social systems there is often a difference between system time as measured by the system events, and clock time which provides a context for those system events. Juggling the interface between system time and clock time is an essential part of planning and management.

When analysing multilevel systems one encounters the intermediate word problem, of finding words to represent parts of the system between the highest level of ‘the system’ and the lowest levels of all the minutiae. Designers and planners come to any problem with a ‘soup’ of vocabulary and ideas based on their previous knowledge and experience. This has to be augmented by observation and, sometimes, the invention of new vocabulary.

In this context we have hypothesised a theory of design in which designers build abstract structures in both top-down and bottom-up ways, until they meet and the design is fully instantiated, with all the once-vague ideas fully grounded in explicit well-defined things.

There are many unsolved problems. Why are so many of our cities alienating and ugly? What does it even mean to say such a thing. Could there be an objective way of measuring the attractiveness or otherwise of a particular cityscape, or an entire city plan? The answer given implicitly here is that cities are structures. If we want to design and plan better cities we must study these structures. No-one would question that cities are multilevel systems. Here we give a way of making this concept well defined and operational. It happens that the representation also has multidimensional structure which is rarely analysed. Surely, to design, plan and better cities one has to be aware of and understand the significance of managing multidimensional events in multilevel systems.

³ At the conference in Ascona we playfully suggested that such representation might reduce the computational complexity to less than the order of N , for example to the order of $N^{1/2}$. We called this the ‘Ascona Conjecture’. Could it be that for some representations, the computational complexity *decreases* with N ?

Acknowledgement

The idea of using simplicial complexes to represent human systems is due to Ron Atkin and emerged during a very creative period of research in the Mathematics Department of Essex University during the nineteen seventies. In particular the idea of structural events is due to Atkin. In those days main-stream research viewed this kind of work as a curiosity on the edge of science. As it turned out, that kind of science was particularly impotent in the face of human complex systems. And thirty years later it seems that the weird ideas rejected by those scientists may be unimaginably potent in understanding complexity and complex systems. Scientists today are independently rediscovering these ideas and applying them to hard problems in the physical, social, and biological sciences. They are a source of endless insights, having a kind of scientific Midas touch. Apply them to any system and you will discover new and interesting things. Structural events indisputably characterise complex system systems, and once you know this the world is a different and more interesting place.

References

- Atkin RH (1974) *Mathematical Structure in human affairs*. Heinemann, London
- Atkin RH (1981) *Multidimensional Man*. Penguin, London
- Barabasi A-L, (2003) *Linked*. Plume Books, Penguin, London
- Gould P, Johnson J, Chapman P (1984) *The structure of television*. Pion Books, London
- Johnson JH (1981) The Q-analysis of road traffic systems. *Environment and Planning B* 8: 141-189
- Johnson JH (1998) Relational structure in planning and design. *Environment and Planning B*, pp 57-67, Anniversary Issue
- Serras J (2005) 'Extending TRANSIMS Technology to an Integrated Multilevel Representation. *Computers in Urban Planning and Urban Management (CUPUM)*, CASA, University College London, UK, June 2005.

The Simulation of Spatial Change: What Relation Between Knowledge and Modeling? A Proposal and Its Application

Silvana Lombardo, Massimiliano Petri

Department of Civil Engineering, University of Pisa, Italy

Abstract

The aim of the research is to investigate land use transformations in a territorial area of Albania, analyzing the connections existing between the deep political and socio-economical changes. The tools we adopted, from the field of the KDS, is able to produce IF/THEN type rules, in which the “IF” part describes the observed state, and the “THEN” part identifies the transition to another state. The application included different phases:

- a) construction of maps referred to different time slices;
- b) construction of automated procedures within G.I.S. in order to perform various kinds of cartographic analysis such as map overlays, neighbouring analysis, etc;
- c) construction of automated report tables referred to cartographic analysis, containing all the attributes necessary to describe the territorial structure;
- d) implementation of an algorithm able to extract land transformation rules;
- e) analysis of the obtained rules in order to find significant relations between socio-economical and spatial evolution.

1 Introduction

The acknowledgment of the complexity of the decision processes has been one of the major problems in the field of spatial planning for over thirty years.

In the last years the concept of a dynamic and processual planning has taken the place of deterministic planning characterized by imposition of rules, zoning, objects, carried out by standards, index, typologies and static constraints and justified by well codified analysis. This new concept of spatial planning is characterized by a multiplicity and a complexity of different actors, objectives, feedback and by the active role of the “time” whether as constructor/destroyer of interests, opportunities, resources or as the dimension of the circularity and learning characteristics of the planning process.

The complexity of the interactions of the different actors who are involved in the decision processes regarding the actions of territorial transformation makes these processes not predictable, not concordant in scope and then lacking of possibility of convergence (Faludi 1987).

But, in the new vision of the operational research and of the urban modelling, a new concept of decision making arises, based on the search of the “optimal compromise” (Roy 1979, 1985; Vinke 1981, 1992) instead of on the search of the “best” solution.

The search of such compromise is located in an evolutionary context in which knowledge stems from the preference systems of the different actors, of possible alternatives and of their effects. In this context decision support tools are developed and these tools, combined with the development of data bases, generate the systems known as Decision Support System, in which a data base is combined with query procedures and at least with one method of comparison and ordering of alternatives (Van Delft and Nijkamp 1977; Bana and Costa 1990; Vincke 1992, Malczewski 1999; Lapucci et al. 2005).

These points of view are based on explicit rejection of deterministic approaches to the decision analysis and are developing quickly, thanks to the great opportunities provided by the powerful computer tools and their abilities to manage very large databases, to execute more complex comparisons and to support the new simulation tools.

In the last decade their fields of applications increased very much: from the decisions in the private companies to the decisions in the public sector to the collective decision, which was studied since the fifties (Arrow 1951; Martelli 1983).

In the field of decision making different authors have developed important contributions to analyse both the technical and the epistemological aspects of the concept of decision and its processes. These contributions lead to the idea that decision making is a process in which the aim of the analysis and of the models is not to identify the “right solution” or the optimal solution but to make explicit the reasons for which an alternative is preferred to another one and the reasons for which a group of actors get more advantages than another one.

The strength of this approach depends on two fundamental components:

- on the one hand, the component which is based on the incremental construction of the decision and on a programme of actions or a set of collective decisions to which the master plan can be assimilate (see, Faludi 1987)
- on the other hand, the component which is based on the construction of knowledge: the ability to open more and more "knowledge windows" about the nature and the evolution of the process of the spatial transformation by means of advanced systems of elaboration, description and simulation.

As to the latter component, the main objective of the tools belonging to the field of spatial modelling and simulation is to open new knowledge windows by working on the relations acting in space and time between spatial phenomena.

2 What Knowledge to Build Knowledge?

A question yet arises: *What knowledge for building knowledge?* How do we select and represent the relations that connect elements in space? This question become

more and more relevant as the availability of computer simulation tools is rapidly increasing. Here, in particular, we refer to spatial simulation tools which, based on the definition of the territorial system elements (agents, sub areas, ...) and of the environment enveloping them, allow the simulation of the working of the relations (evaluation and choice rules) between elements and between elements and their environment (e.g. Cellular Automata, Multi Agent Systems, ...). In this context a fundamental issue to be considered is: given the elements and the environment of interest, how such relations are defined?

In the simulation applied to territorial sciences, the building of evolution rules applied to the “actors” (be they individuals, groups, or spatial unities) follows three main approaches:

1. the adoption of well established theories: in it, the rules derive from analytical models (e.g. the Harrys and Wilson’s model (Wilson 1981, 2000; Lombardo and Rabino 1983) or economical models such as Discrete Choice Models (Heppenstall et al. 2003);
2. the “building” of rules based on “expert knowledge”: the rules derive from the know-how of the modeler or from common sense. In this case predefined rules are used that can follow the BDI framework (Beliefs, Desires and Intentions) (Nonas and Poulouvassilis 1998) or the “Activity Based Approach” (Zwerts 2004) or rules used in “Active-Databases”, designed mainly to build reactive agents;
3. The “discovery” of rules: this is the more complex: the rules derive directly from data, by means of techniques of Knowledge Discovery in Databases, belonging to the field of Artificial Intelligence, such as Decision Tree Induction, Neural network, Genetic Algorithms. This way often produces limited sets of rules, because it needs still more studies and developments, but presents some advantages, as it allows to:
 - extract (possible new) knowledge from very large spatial data bases
 - extract possible rules specific for the study area
 - verify the rules adopted in 1) and 2)
 - verify if and what rules are unchanging in time
 - build new theories

2.1 The Theory Based Approach

As to the first approach we find some hybrid models that use the interaction between the meso-macro scale dynamics and the micro scale one; an example is the work of Heppenstall et al. (2003) where a spatial interaction model is used to simulate the oil price market; the agents, that are the petrol pumps, obey to the following rules:

- if the profit grows, continue with the actual strategy;
- if the profit decreases, change strategy (increase the price or reduce it);
- if the profit is uniform, keep the price constant.

In other works the approach of dynamic spatial interaction is used, as in IMADYN (Interactive Multi Agent Dynamics), which implements an interactive MAS in order to simulate the location dynamics of market economic activities in an urban area in a changing environment. After some prototypical experiments (Lombardo et al. 2003), it was applied to the Pisa-Lucca-Livorno urban system (Lombardo et al. 2004).

Furthermore, there are the models based on statistical and economical theories like UrbanSim (Waddell et al. 2003), one of the few Multi Agent Systems really operational in the cities of Honolulu, Hawaii and Salt Lake City. In this modular model the agents are the workers, the households and the goods owners, which interact following a discrete choice model, in particular a multinomial logit model (Ben-Akiva and Lerman 1985), the process produced the location of population, work places and other. Moreover, in it a statistical regression model is used to forecast land price evolution.

Another model, simulating the dynamics and the agents behaviour by means of rules based on analytical formulations, was implemented by White and Engelen (White and Engelen 2000). It is a Cellular Automata where land use evolution is simulated by means of transition rules expressed by a mathematical formula; this formula gives a potential change value to each cell of the cellular lattice and is a function of the neighbourhood effect, of the accessibility and of the land use class suitability.

2.2 The A-Priori Knowledge Approach

The second approach is at present the most used and includes models aimed to simulate the pedestrian movement inside the city (Hanisch and Schultze 2003; Schelhorn and O'Sullivan 1999), models related to the growth and development of the city (Semboloni et al. 2004) or models which simulate the evolution of a rural society or other simplified environments (Cioffi-Revilla and Gotts 2002; Axtell et al. 2002).

An element to be stressed is the weight of the rules: a small difference in the behavioural rules can lead to totally different dynamics (Otter et al. 2001). More in general, two different rule formulations both theoretically and conceptually acceptable can produce different results.

This implies the need to create a meta-model of the agent behavioural rules scientifically sound and including a "decision tree" framework regarding the process implied in each mental map (Occelli 2004). In this context, it is to be noted the important contribution given, among others, by psychology and sociology, with the introduction of the "Social Network" concept (Jager and Janssen 2003). These agents relationship framework introduces some psychological variables such as the social relations, the preferences and the individual needs in the market dynamics. The utility of buying an item depends on the individual preferences, on the information received from friends and on the weight given to the social relationships by the single agent. Moreover, uncertainty can lead to imitative behaviours.

Following the previous hypotheses, agents are splitted on the basis of two variables: satisfaction need (social and personal) and uncertainty level. By transferring these variables in mathematical form, it is possible to simulate the agent's cognitive process.

In recent years the simulations belonging to the "Activity-Based Approach" are increasing (Arentze and Timmermans 2000; Tabak et al. 2004); it allows to build micro-scale urban simulations by means of individual activity daily diary. It appears interesting, among other things, because it support the selection of interventions aimed to influence individual participation to urban activities and transport demand; examples of such interventions are changes in the shops opening time, in activity locations or in the land use development.

An alternative methodology to the Activity-Based is the approach based on the BDI (Beliefs, Desires, Intentions) framework, (Nonas and Poulouvassilis 1998), for some aspects similar to the "Active-Databases", where links between the classes of the Database are built in a Database Management System (DBMS).

In this case, an object/agent is informed of an event (for example, a new residential area or a new road) thanks to the relations built between class instances; such relations, which allow to exchange messages between the instances themselves, under certain circumstances can cause the agent reaction. In both cases the rules (in the form of conditional IF-THEN rules) are predetermined. It is useful mention that in the "Active-Database" the UML (Unified Modeling Language), an event-driven language, can be used to link the objects of each class and to implement the rules.

In their work, Nonas and Poulouvassilis modify the rules applied by each agent, thanks to the use of a Genetic Algorithm that selects, at each step, the best rules in comparison with an optimization function.

Once again however it is used a predefined set of rules and of targets in which the Genetic Algorithm makes its optimal choice.

2.3 The Knowledge Discovery Approach

In territorial planning process, the quality of the evaluation procedure increases with the knowledge of the dynamic behaviour of the spatial system for which decisions are requested. The factors and relations that influence such behaviour have been theorised, formalised and modelled since the birth of urban and territorial studies in many disciplines. During the last few years, as said before, such theories and models (including the "dynamic revolution" of the eighties) have been based on the explicit formulation of the rules, the relations and the forces which trigger the actors choices and then determine both city's and territory's evolution. But, if we suppose that nowadays some of such rules and relations could have changed, we need to develop a strongly aimed theoretic and operational field of study: knowledge building and management in territorial planning process.

For this purpose it is therefore necessary, in our opinion, to select and implement innovative tools able to handle the huge amount of available data concerning

territorial systems (statistical, economic, cartographic data and so on) in order to extract useful information from them.

In this context, we propose the building of cognitive systems (intelligent systems) which, observing a given reality, be able to extract knowledge on the role played by urban/territorial factors and by their relations in urban/territorial evolution and change. We describe goals and tools of the “knowledge builder and manager”, and illustrate the results of the applications of some of these tools to the investigation on the roots of land use change in an urban-agricultural dynamic context.

On the basis of a survey on the state of art about the presently available tools, we selected some derived from Artificial Intelligence and from the field of G.I.S. Though the elaboration of various kind of available data (thematic maps, population density, transport infrastructures, productive settlements, maps of archaeological and historical heritage, etc.) we are able to extract and build knowledge directly from experimental data and it is also possible to represent the extracted knowledge in a very efficacious and communicable way, in the form of sets of spatial transformation rules. Afterwards, these rules could be analyzed and compared with traditional rules we adopted, until now, to explicate urban/ territorial dynamics, in order to change/ correct some theoretical issues. For this purpose is necessary to dispose of large amounts of data: however, the value of such data mainly depends on our ability to exploit knowledge from them.

In the following paragraph we describe one of the experiments we carried out in this direction, treating in detail both the characteristic of the adopted dataset and the methodology we adopted for the extraction of the rules.

3 The Application

The case study of the application here described is related to the territory of the Municipality of Preza, in Albania. The aim is to investigate the land use change that took place in the periods between 1991- 1996 and 1996- 2003.

This is an area where the deep political socio-economic changes occurred make necessary to verify the feasibility of future land use development.

3.1 Methodology and Tools

The study of the described phenomena was carried out by using and integrating each other two kinds of tools: G.I.S. and automatic learning tools (Bonchi et al. 2004a, 2004b).

In the first phase, by means of a G.I.S. software, the geographical georeferenced land use database related to three different years was built. By means of spatial elaborations performed within the G.I.S., the information derived from the various maps are put in the form of relational tables, where in the lines are reported the elementary cells through which the territory has been represented and in

the columns the attributes that give information about the characterization of each single cell.

In the second phase the table is given as input data of the automatic learning tool, able to perform the extraction of a set of land use transformation rules.

In the third phase, we applied the extracted rules to the thematic map related to the first temporal section, and we verify, by means of the G.I.S. tool, the reliability of these rules, by comparing the structure of the territorial system at the second temporal section with the one obtained by applying such rules.

3.1.1 G.I.S.

The first phase was related to data consolidation and preprocessing, performed within G.I.S., as it makes possible both to store a remarkable amount of digital cartographic data by georeferencing them with respect to a same reference system, both to perform on them any kind of spatial analysis.

The study area was modelled as a vectorial grid containing over 10.000 squared cells (50 x 50 m), able to memorize many attributes for each cell of Preza municipality. On the basis of three land use maps and some vectorial spatial data, thematic maps were built, by using Overlay Mapping and Geoprocessing tools.

The information coming from land use maps had to be integrated with other information related to the presence of anthropic infrastructures, social and demographic data and morphologic variables, assigned to each cell. At this point it was possible to select the elements in the neighbourhood of every elementary cell. The characterization of "neighbourhood" was performed by means of Model Builder within ArcGis 9 Software, able to create a very efficient data analysis computational model, even in presence of large amount of data. Figure 1 illustrates the model built for the neighbourhood analysis related to single land uses. It allows to extract automatically, for each cell, the number of cells of each land use located within its neighbourhood, consisting in 5 x 5 cells, namely a square with each side measuring 250 meters.

In Fig. 2, an instance is represented where the "core cell" presents a neighbourhood consisting in 14 cells of Land use number 8, 7 cells of Land use number 5 and 4 cells of land use number 9.

The results of these elaborations are put, by using automated procedures, in the form of relational tables in which, for each spatial entity, there is a description (defined by means of alphanumeric attributes) of the state of the cell and a description of the states of the neighbouring entities. Such table represents the input data (training set) for the automatic learning tools that allows the explanation of the relationships among the elements of the studied area, by generating a set of transformation rules.

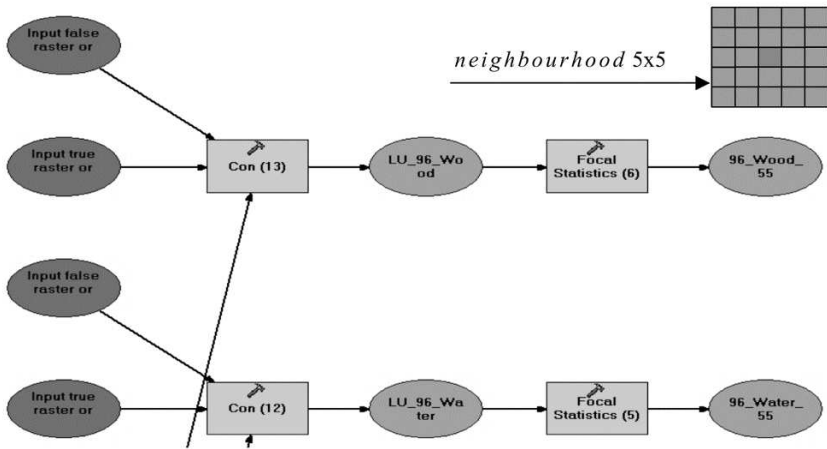


Fig. 1. A section of the model used for neighbourhood analysis

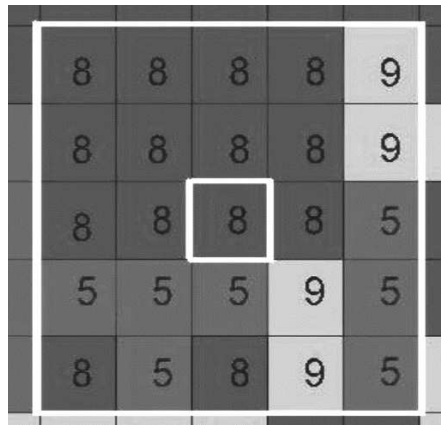


Fig. 2. Characterization of the neighbourhood of each cell

3.1.2 Knowledge Discovery Systems

In the last few years the problems related to handle increasing amount of available data (in particular, territorial data) has produced a growing interest with regard to tools able to automate learning processes. In this context, a new technology emerged, named “Knowledge Discovery in Databases” (or KDD in short), consisting in a number of systems and tools oriented to discover useful and interesting knowledge from data in large databases. Among them, we selected a classification

method, able to perform learning from example, namely the classification based on Decision Trees Induction (DTI).

DTI is a supervised classification technique: initial data consist in a certain number of records each of them presenting multiple attributes (i.e: land use at the first year, activity density, slopes, etc.) and a class label, that is the classification target attribute (in our case, land use at the second year).

In classification, the records constituting the training set are analysed in order to produce an accurate description or model for each class: this model can be applied later in order to classify further and/or future cases whose classes are unknown. In our case, records are related to each single cell within a “vectorial grid” and contain a series of attributes. In this application, we used the algorithm C4.5.

The classification results can be visualised by using a tree structure like a flow chart in which we can find a root, edges, nodes and leaves (see Fig. 3).

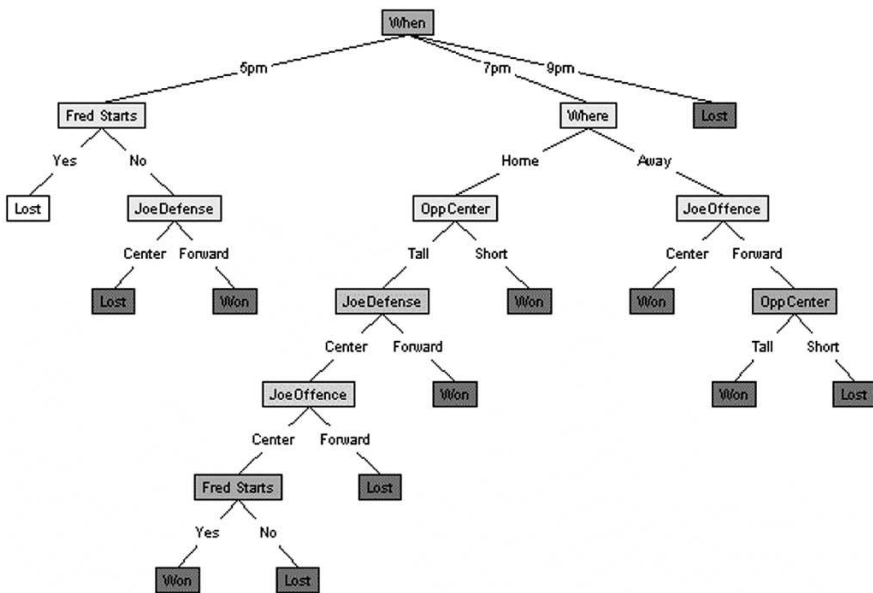


Fig. 3. An example of Decision Tree Induction

The root represents the attribute on which the partition was performed, the nodes are labelled with the names of the remaining attributes, the edges with the possible values that the attributes can assume and the leaves are labelled with the different classes. Such framework is easy to be interpreted, as in it each path going from the root to one leaf, through the edges corresponding to the attribute values, represents a classification rule. It is therefore possible to classify cases for which label class is unknown, using decision trees as an estimating technique. Each rule presents a value of “support” and “confidence”, able to measure frequency and

strength of the rule. Rules that are valid for many cells present an high value of support, while very precise rules present high value of confidence.

4 Data and Results

The general assumption of our experimental work is that the phenomena lying at the basis of territorial change are:

- the land use and the presence of various kinds of infrastructures or morphologic characteristics in each elementary territorial cell;
- the spatial relations of each cell with other cells, measured in term of ‘enlarged adjacency’.

For each spatial entity there is then a description (defined by means of alphanumeric attributes) of the state of the entity itself and a description of the states of the neighbouring entities.

Moreover, before running the algorithm, some preliminary statistical analysis were performed by using OLAP (On-Line Analytical Process) Cube, in order to have some previous information useful to better set parameters during the mining phase (see Fig. 4).

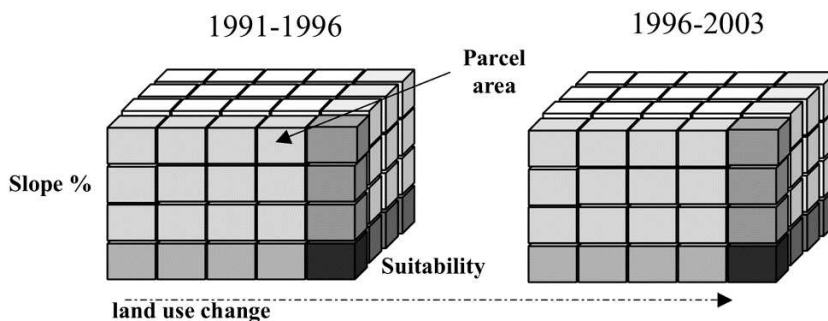


Fig. 4. The 4 dimensions of the OLAP CUBE

4.1 Input Data

The input data for this application were provided by LUP2 (Land Use Project), developed in Albania. The geo-database developed within the LUP2 project was designed for a central agency or State organization with the aim of becoming the central node for the Albanian Information System; the hosting agency is the Soil Research Institute (SRI), on behalf of the Ministry of Food and Agriculture.

4.2 Characterization of the Variables

An international agreement on land-use classification does not exist. There is a significant diversity of opinion about what constitutes a land-use (UNEP/FAO 1994). The term land-use has different meanings across disciplines and, as a result, implies a set of mostly unidentified characteristics.

The first phase of our work was therefore aimed both to homogenize the different land use legends related to the three maps and to find an accurate and representative set of land use classes, able to characterize the whole territory of Preza Municipality.

As to the homogenization, we choose to refer to Luisa legend (Land Use Information System of Albania), related to 2003 land use map: all data have been imported in a GeoDatabase, by means of Case Tools (Computer Aided Software Environment) and Visio software, so that it was possible to perform links/ relations between tables related to different land use legends.

As far as the second task is concerned, starting from the land use definition described above, rules and criteria were applied for the development of a final classification. The main distinctive criteria adopted for the description and classification of the various land-uses were: (1) function and, subsequently, (2) activity. Function refers to the economic purpose of land-use and can group many different land-use types in a single category; activity is defined as “the combinations of actions that result in a certain type of product” (UN 1989) and refers to a process.

On the basis of these criteria land-use resulting classes were systematically structured in a hierarchical order to be able to regroup them at different levels.

The resulting classes are presented in Table 1.

Table 1. Land uses classe

Crop	Service
Fruit Tree	Transport
Industrial area	Uncultivated
Meadow/pasture	Unknown
Olive-grove	Vineyard
Recreation	Water
Residence	Wood

The second phase was aimed to find variables able to describe the land use changes occurred in that area during the considered time periods. In general, possible land-uses are limited by biophysical constraints that include climate, topography, soils and geological substrate, presence or availability of water and the type of vegetation.

Agricultural practices differ from one region to another and different types of cultivations are practised on the same type of land in different areas, depending on history, local traditions and way of life, besides the biophysical constraints. The

location of an area with respect to other types of land-uses, such as nearby residential and industrial areas, is also an important factor (e.g., the location of Preza municipality close to main urban centres like Tirana and Durrës as well as its proximity to Rinas airport) and economic incentives as part of policy (e.g., as well as the EU Common Agricultural Policy) has also a great influence on the land-use patterns. In this context we selected a set of variables able to affect the land use change occurred in that period, considering both functional and physical aspects.

The choice of variables also depends on initial data availability and from the observation of some land use dynamics. In fact, as we realized that the dimension of land use parcels strongly influences their transformations, we chose to include as initial variable the original area of the parcel in which the examined cell is included. For the same reason the very different behaviour between edge and internal regions of the same land use parcel suggested to introduce the distance of each cell from the border of the parcel containing it.

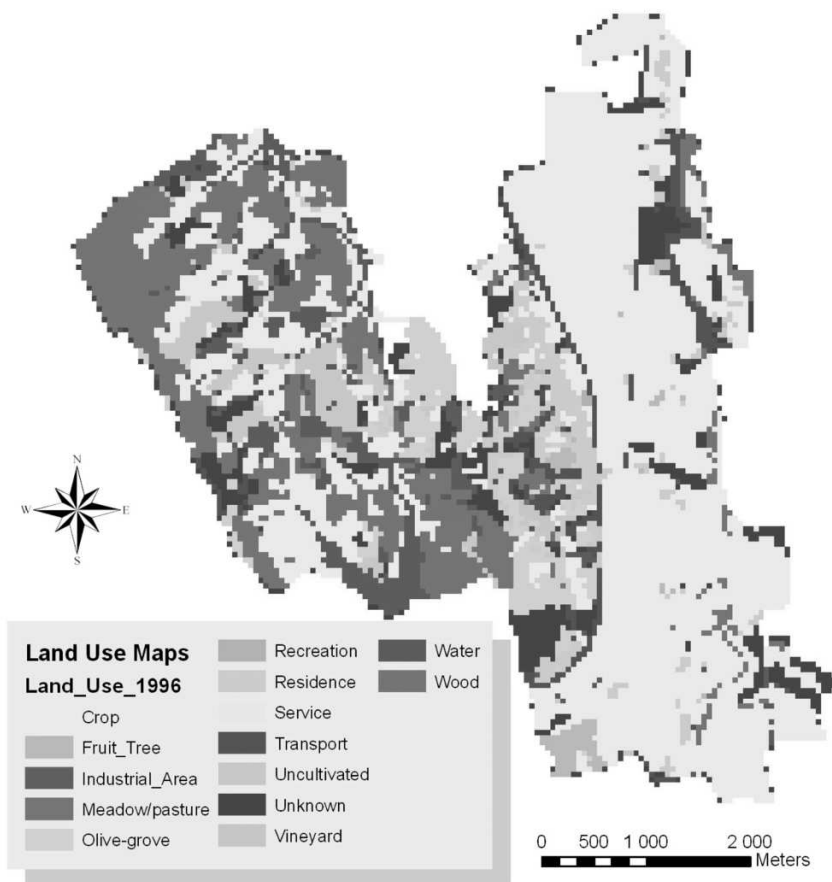


Fig. 5. Land use map related to 1996

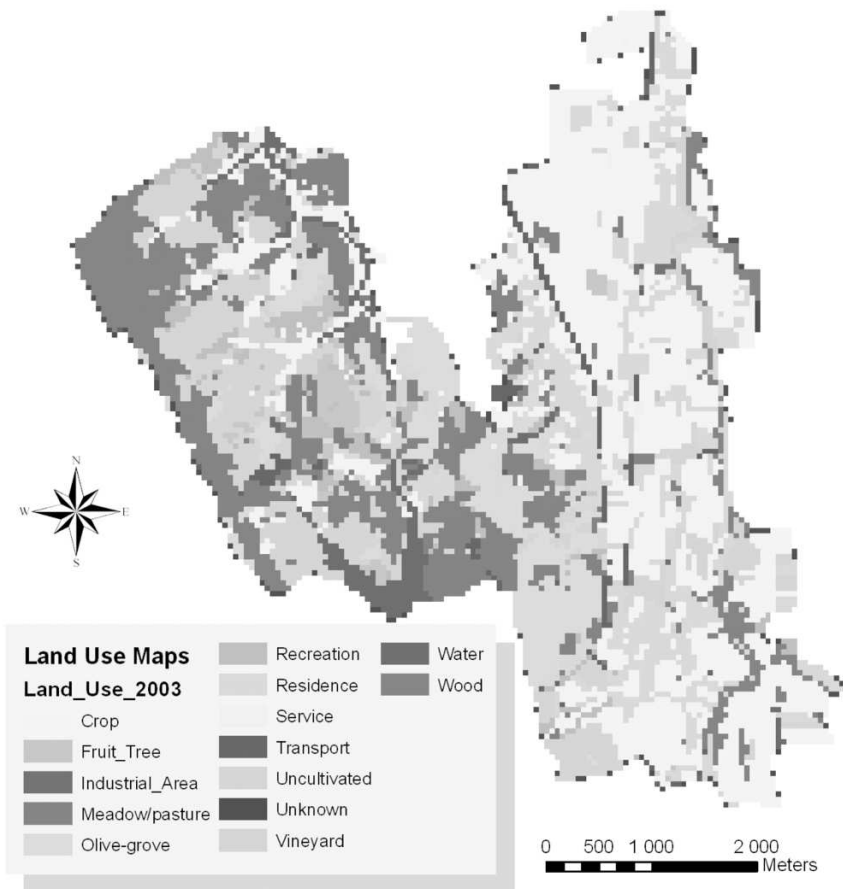


Fig. 6. Land use map related to 2003

Figures 5 and 6 represent the homogenized land use maps we used for building some attributes of the relational tables, as mentioned in 1.1.

4.3 Data Analysis

Beginning from the results of multitemporal analysis (see following paragraph) we realized that DTI is not able, in the specific case of Preza, to analyze at the same time tree different land use maps, because the two couple of maps (1991-96 and 1996-2003) are representative of two periods presenting very different dynamics. In fact while the first idea was to use the extracted rules in relation to 1991-1996 period as training set for the algorithm and then validate these rules with the maps related to the second period (1996-2003), the results of the multitemporal analysis

showed that the phenomenon of land fragmentation occurred in 1992 (but effective only from 1996) makes very different the land use dynamics of the two periods, so that we held the rules related to 1991-1996 not able to explain the transformations occurred in the second time slot.

Moreover, a more detailed study concerning the “new” territorial rules (emerged after land fragmentation phenomenon) was performed by means of DTI. In the next paragraph we present the results of these tests.

4.4 Multitemporal Data Analysis

For the preliminary multitemporal statistical analysis we used OLAP (On-Line Analytical Process) Cube, a Datawarehouse technique, considering four initial variables, namely

- Land use change class
- Analysis time section of (1991/96, 1996/03)
- Land slope class
- Land suitability (as assessed by SVALTEC s.r.l.)

This first analysis shows the very different evolutions between the two periods. An historical analysis related to land property in Albania in this period allows to better explain the reasons of these differences.

In 1992, as a consequence of the deep political changes occurred in Albania, there was a total reallocation of land that caused a fragmentation of original cadastral parcels into small parts in consequence of the transformation of land property from public to private: the State decided that a portion of 0,7 hectares had to be equitably distributed to each person, as the documents certifying the old land property were not found. However, until 1996, free transaction of lands was forbidden: it is therefore evident that until 1996 land use dynamic has been paralyzed, while after this date we assist to a deep transformation of the territorial structure.

In this context, the results of this first OLAP analysis show that:

- In the period 1991/1996 there is a certain stability as far as land use is concerned, with about 39% of correspondence and medium level land use changes homogeneously distributed within the territory as to various slopes and “Land Suitability” classes.
- In the same period transformations are uniformly distributed between the different land use classes and slope categories. Moreover there aren’t significant conversions of land use but only some medium level modifications.
- In the period between 1996/2003, differently from what happened before, portions of land presenting high values of slope have been abandoned (15%); it was probably because during the dictatorship, steep land management by means of terracing was imposed.
- Moreover, land private management caused a reduction of the woods, encroached by agriculture: in this period the transformation from wood to

meadow and pasture amount to 9,9%. As expected, there is a strong correspondence between slope classes and land use classes: steep lands are always related to less profitable land use classes, such as wood and meadow and pasture, and vice versa.

4.5 Mining Phase: Extraction of Rules

For analysing the time slot 1996-2003, we used 29 variables as input data for the decision tree (Table 2).

The prediction model provided rather good results, as it presented a land use prediction accuracy of about 80%. Moreover, correlation between real and simulated values at 2003 is very good, equal to 0,75 and a very low difference between the square root of the medium quadratic classification error (0,149) and absolute medium error (0,0444) shows that there are no classification outliers, that is to say predictions too far from medium error. Also prediction accuracy for each land use class is very good for almost all classes ($>0,7$), except for Service, as it presents a very small number of cells in 1996, so that the rules related to these classes are difficult to test and evaluate. Confusion matrix provides information on error distribution: errors are mostly associated to classification of Crop land and wood, as these classes contain the higher number of cells, namely 3363 and 1261.

Table 2. Tables of attributes for each cell

Attribute	Description	Domain values
SLOPE	Slope class	(1-4)
WATER_NAT	Distance from natural water sources	(N; B; MB; M; MA; A) ¹
IF_EDIFIC	Presence of building	(0,1)
NUM_EDIF	Density of building	(M, N, P, T) ²
AREA_PART	Area of original (1996) cadastral parcel	(N; B; MB; M; MA; A) ¹
EDIF_R500	Density of buildings within 500 meters	(N; B; MB; M; MA; A) ¹
ROAD_DIST	Distance of nearest road	(1-7)
ROAD_SURF	Surface type of nearest road	(0-3) ⁴
ROAD_COND	Maintenance condition of nearest road	(0-3) ⁵
ROAD_TYPE	Classification of nearest road	(2-5) ⁶
WATER_ART	Distance from artificial watering canal	(N; B; MB; M; MA; A) ¹
EROS_RISK	Erosion risk	(1-4)

Table 2. Continued

Attribute	Description	Domain values
DRAIN_VAL	Value of land drainage class	(1-4)
LANDUSE_96 LANDUSE_03	Prevalent land use class	(C, Wa, Wo, V, U, U, T, S, F, Rc, M, Rs, O) ³
DISTBOUN96	Distance from the edge of cells in 1996	(1-7)
<LandUse>_55	Number of cells of each land use located within a neighbourhood with sides of 250 meters, centred on examined cell	(N; B; MB; M; MA; A) ¹

¹ (N= nothing; B= little; MB= medium-little; M=medium; MA= medium-high; A=high)

² (M= medium; N= nothing; P= little; T= high)

³ (C= Crop, Wa= Water, Wo= Wood, V= Vineyard, Uk= Unknown , Uc= Uncultivated , T= Transport, S= Service, F=Fruit Tree, Rc= Recreation, M= Meadow/pasture, Rs= Residence , O= Olive-grove, I= Industrial area)

⁴ (0=N.A.; 1=Paved; 2=Unpaved, 3=gravel loose)

⁵ (0=N.A.; 1=Good; 2=Fair, 3=Poor)

⁶ (2=Main road, 3=Secondary road, 4=Urban main road, 5=Urban secondary road)

4.6 Results: Emerging Territorial Dynamics in the “New” Albania

In this first application, more over than 300 land transformation rules have been extracted, in the period between 1996 and 2003. In this chapter, some of the most representative ones are described. We built a more aggregated explanatory level to express the extracted knowledge in a more synthetic and communicable way: we divided extracted rules in two groups, namely transformation and inertial rules, giving for each group the conditions that influenced the evolution or the preservation of the original land use.

In the following tables we also report the support (number of cells involved) of each rule.

These rules were then implemented in a Cellular Automata using the software NetLogo as already done in the case study of Pisa (Lombardo and Petri 2005). In this case study, we used a deterministic approach to the rules application. Afterwards, we realized that in such a way, being the results of DTI not deterministic, a significant part of the obtained information (that’s to say the variance of the dataset) went lost.

Table 3. Transformation rules

'96	Main Conditions		'03	N°
WOOD	Good road condition; Original parcel medium-wide size;		MEADOW	95
	Distant from artificial watering canal;			44
	Low drainage class; Original parcel small size; Unpaved roads			26
CROP	High slope; Poor road condition; Medium erosion risk;	Low number of buildings within 500 m.; Urban secondary roads;	UNCULTIVATED	78
		Original parcel medium size;		73
	High slope; Medium erosion risk; Low number of buildings within 500 m.;			289
	Low erosion risk; Poor road condition; Distant from artificial watering canal;			61
CROP	Very low slope;		RESIDENCE	182
	Low erosion risk; Original parcel medium- wide size; Medium road condition;			118
	Low erosion risk; Original parcel medium-wide size; No water in neighbourhood; Natural watering canal inside;	Fair road condition; No residence in neighbourhood; Full of crop in neighbourhood;		49
		No transport in neighbourhood;		69
WATER	Low number of buildings within 500 m.; Distant from artificial watering canal;		RECREATION	40
	Unpaved road; Original parcel small size;			62
	Low erosion risk;	Original parcel medium-small size; Medium distance from natural watering canal;		42
		Paved roads;		64

Table 4. Inertial rules

'96- '03	Main Conditions		N°
WOOD	Low-medium number of buildings in 500 m.; No residence in the neighbourhood;		147
	Original parcel medium size;		134
	Original parcel small-medium size;		50
	Original parcel very large size;		502
	Paved roads;		41
CROP	Very low slope; Original parcel small size; Medium distance to roads;	No additional condition;	136
		Low drainage class;	148
		Poor road condition; Low number of buildings in 500 m.; Unpaved roads; No water in the neighbourhood; No transport in the neighbourhood;	109
	Unpaved roads; Low-medium number of buildings in 500 m.; Roads inside cell;		64
	Low erosion risk; Original parcel small size; Very low slope	Medium distance to roads	229
		Low distance to roads Medium drainage class Low- medium number of buildings within 500 m.	151
	Low erosion risk; Urban secondary roads;	Near to artificial watering canal; No water in the neighbourhood; Original parcel small-medium size;	73
		Artificial watering canal inside; Medium drainage class; Original parcel small-medium size;	51
		Fair road condition;	92
	Very low slope; Original parcel small size; Roads inside cell; Urban secondary roads; Unpaved roads; No residence in the neighbourhood;		643
	Low erosion risk; Original parcel small size; Urban secondary roads;	Paved roads;	60
		Low drainage class; No residence in the neighbourhood	131

Table 4. Continued

'96- '03	Main Conditions		N°	
CROP	Low erosion risk Original parcel small size; Urban secondary roads;	Poor road condition; No water in the neighbourhood; No residence in the neighbourhood High drainage class;	86	
		RESIDENCE	Gravel/lose roads;	325
		OLIVEGROVE	Low number of buildings in 500 m.;	260
MEADOW	Paved roads;	183		
UNCULTIVA- TED	Low number of buildings within 500 m	Low distance to roads; Unpaved roads;	54	
		Medium distance to roads; Unpaved roads;	64	
		Little water in the neighbour- hood; No Meadow in the neighbour- hood	48	

Therefore, in this application we did not use deterministic rules but, to maintain the residual variance of the training dataset, we implemented probabilistic assignment rules (Arentze and Timmemrmans 2000); the probability of selecting the THEN response, also in the case of true value of the IF statement, is for the moment given by the relation between the number of training cases at the actual leaf of the tree correctly classified and the total number of cases classified in that node.

5 Conclusions

The research work carried out and the study case developed show, in our opinion, that the most promising way in the direction of building behavioural rules which drive agent-based micro-scale simulations is the use of Data Mining tools, such as Genetic algorithms, Neural Network, Decision Tree Induction. This last tool turns out to be specially suited for territorial/social analyses, as it produces rules easily interpretable and then allows both to individuate specific local dynamics and to provide models which are simpler and transparent for evaluation and decision purposes. This property cannot be found in complex mathematical models, often based on equations with parameters to be estimated, difficult to be calibrated and whose physical meaning often is lost.

Moreover, it is important to integrate micro-scale models, based on rules extracted by means of Datamining techniques, with meso scale models, as those used

in the application described in 2.1. This could lead to the revision of classical theories.

Moreover, such integrated tool will allow to solve the border problems of the micro-model, by working on a system which is not closed, but interacting with neighbouring areas through behavioural meso scale rules (see Fig 7).

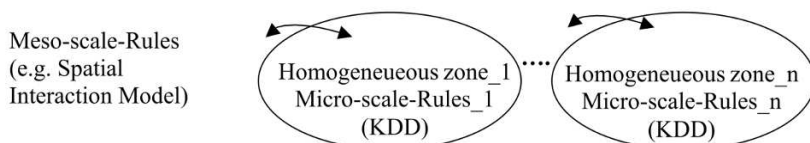


Fig. 7. Scheme of a hybrid model

Acknowledgements

We thank the Svaltec s.r.l. firm of Florence for the kind authorization to the use of Dataset.

References

- Agrotec S.p.A. Consortium (2003) Land Use Planning Methodology. Technical Report 8. EU Phare Land Use Policy II Project, Agrotec S.p.A., Rome
- Agrotec S.p.A. Consortium, 2003. Preliminary Land Use Plan for Preza Commune. Technical Report 17. EU Phare Land Use Policy II Project, Agrotec S.p.A., Rome
- Arentze TA, Timmemmans HJP (2000) ALBATROSS – A learning-based transportation oriented simulation system. Urban Planning Group/EIRASS, Eindhoven University of Technology, The Netherlands
- Arrow K (1951) Social Choice and Individual Values. Cowles Foundations and Wiley, New York
- Axtell RL, Epstein JM, Dean JS, Gunerman G, Swedlund A (2002) Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. Proceedings of the National Academy of Sciences of the United States of America **99**: 3: 7275-7279
- Bana e Costa C (1990) Readings in multiple criteria decision aid. Springer, Berlin
- Ben-Akiva M, Lerman SR (1985) Discrete Choice Analysis Theory and Application to travel demand. The MIT Press, Cambridge, MA
- Bonchi F, Lombardo S, Pecori S, Santucci A (2004a) Learning about Land Use change in Rome and in Pisa Urban areas. In: Diappi L (ed) Evolving Cities: Geocomputation in Territorial Planning. Ashgate, England

- Bonchi F, Lombardo S, Pecori S (2004b) Knowledge discovery and data mining to investigate urban and territorial evolution: tools and methodologies. In: Diappi L (ed) *Evolving Cities: Geocomputation in Territorial Planning*. Ashgate, England
- Burrough PA, McDonnell RA (1997) *Principles of Geographical Information System*. Oxford University Press, Oxford
- Cioffi-Revilla C, Gotts NM (2002) Comparative analysis of agent-based social simulations: GeoSim and FEARLUS models. Paper presented at the Model-2-Model Workshop on Comparing Multi-Agent Based Simulation Models, March 31-April 1 2003, Marseille, France
- Faludi A (1987) *A Decision Centered View of Planning*. Pion, London
- Fayyad U, Piatetsky- Shapiro G, Smith P, Uthurusamy R (1996) *Advances in knowledge Discovery and Data Mining*. AAAI/MIT Press, Mento Park, CA
- Han J, Kamber M (2000) *Data mining: concepts and techniques*. Morgan Kaufmann, S.Francisco
- Hanisch A, Schultze T (2003) Online simulation of pedestrian, flow in public buildings. In: Chick S, Sanchez D, Ferrin D, Morrice DJ (eds) *Proceedings of Winter Simulation Conference*, Association for Computing Machinery, New York
- Heppenstall AJ, Evans AJ, Birkin M (2003) A hybrid multi-agent/spatial interaction model system for petrol price setting. *Transaction in GIS* 9: 1: 35-51
- Holsheimer M, Siebes A (1994) *Data mining: the search for knowledge in databases*. CWI Technical Report CS-R9406, Amsterdam
- Jager W, Janssen M (2003) The need for development of behaviourally realistic agents. In: Sichman JS, Bousquet F, Davidsson P (eds) *Multi-Agent based simulation II*. Lecture Notes in Artificial Intelligence, Springer, Berlin, pp 36-49
- Lapucci A, Lombardo S, Petri M, Santucci A (2005) A KDD based multicriteria decision making model for fire risk evaluation. *Proceedings of 8th AGILE 2005-Conference on GIScience*, 25-29 May, Estoril, Portugal
- Lombardo S, Pecori S (2004) Investigating Urban structure and evolution through Artificial Intelligence tools. *Studies in Regional & Urban Planning*, Special Number, pp 32-52
- Lombardo S, Petri M (2005) Un sistema di supporto alle decisioni: integrazione di analisi GIS e Datamining. In: Las Casas G, Pontrandolfi P, Murgente B (eds) *Proceedings of the conference URBING 2005*, 29th-30th April, Potenza, Italy, pp 15-27
- Lombardo S, Petri M, Zotta D (2003) Urban services spatial dynamics in a changing environment simulated by a multi agent system. *13th European Colloquium on Theoretical and Quantitative Geography*, vol 1, Lucca, Italy
- Lombardo S, Petri M, Zotta D (2004) Intelligent GIS and retail location dynamics. *International Journal Computational Science and Its Application (ICCSA 2004)*, Assisi, Italy, vol 2, pp 1046-1056
- Lombardo S, Rabino GA (1983) Some simulations on a Central Place Theory Model. *Sistemi urbani* 5: 2: 315-332
- Malczewski J (1999) *GIS and Multicriteria Decision Analysis*. John Wiley, New York
- Martelli P (1983) *La logica della scelta collettiva*. Il Saggiatore, Milano
- Nonas E, Poulouvassilis A (1998) Optimisation of Active Rule Agents Using a Genetic Algorithm Approach. *DEXA*, pp 332-341
- Occelli S (2004) A perspective on MAS approach in urban modelling, IRES-Istituto di Ricerche Economico Sociali del Piemonte

- Otter HS, van der Veen A, de Vriend HJ (2001) ABLOoM: Location behaviour, spatial patterns, and agent-based modelling. *Journal of Artificial Societies and Social Simulation*, vol 4, nr 4
- Pecori S, Santucci A (2002) The competition for land between urban and rural systems investigated by G.I.S. and Data Mining techniques. *Proceedings of the 3rd International Conference of Decision Making in Urban and Civil Engineering*, 6 – 8 November, London
- Piatetsky-Shapiro G, Frawley WJ (1991) *Knowledge Discovery in Databases*. AAAI/MIT Press
- Quinlan JR (1986) Induction of decision trees. *Machine Learning* 1: 81-106
- Roy B (1979) De-quelle decision s'agit-il, qui aider et comment?. Document du LAMSADE nr. 4
- Roy B (1985) *Méthodologie Multicritère d' Aide à la Décision*. Ed. Economica, Paris
- Schelhorn T, O'Sullivan D, Haklay M, Thurstain-Goodwin M (1999) STREET: an agent-based pedestrian model. *Centre for Advanced Spatial Analysis, University College London*, Paper 9
- Semboloni F, Assfalg J, Armeni S, Gianassi R, Marsoni F (2004) CityDev, an interactive multi-agents model on the web. *Computers environment and urban system* 4: 45-64
- Tabak V, de Vries B, Dijkstra J (2004) User Behaviour Modelling - Applied in the context of space utilisation. In: von Leeuwen JP, Timmermans HJ (eds) *Development in Design and Decision Support System in Architecture and Urban Planning*, Eindhoven University of Technology, pp 141-156
- Van Delft A, Nijkamp P (1977) *Multicriteria Analysis and Regional Decision Making*. Martinus Nijnhof, Leiden
- Vincke Ph (1981) Multicriteria Analysis: survey and new directions. *European Journal of Operation Research* 8: 207-218.
- Vincke Ph (1992) *Multicriteria Decision Aid*. John Wiley, Chichester
- Waddel P, Borning A, Noth M, Freier N, Becker M (2003) *Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim*. University of Washington, Seattle
- White R, Engelen G (2000) High-resolution integrated modelling of the spatial dynamics of urban and regional systems, Pergamon press, *Computers. Environment and Urban Systems* 24: 383-400
- Wilson AG (1981) *Catastrophe theory and bifurcation: applications to urban and regional systems*. Croom Helm, London, University of California Press, Berkeley.
- Wilson AG (2000) *Complex Spatial System: The modelling foundations of urban and regional analysis*. Pearson Education, Harlow, pp 62-77
- Witten IH, Frank E (2000) *WEKA-Machine Learning Algorithms in Java*. Morgan Kaufmann, S. Francisco

A Structural-Cognitive Approach to Urban Simulation Models

Juval Portugali

Environmental Simulation Laboratory (ESLab), Tel Aviv University, Israel

Abstract

This paper focuses on two interrelated properties of the agent-base (AB) and cellular automata (CA) urban simulation models currently employed in the study cities: The global urban structures that emerge out of their dynamics plays no role in the dynamics itself and, agents' behavior as postulated in these models disregards the basic principles of human cognition, behavior and action as revealed by cognitive science. These two properties are interrelated because empirical and theoretical studies in cognitive science indicate that agents' behavior in cities is strongly influenced by the global structure of cities. The paper elaborates on the deep roots of these properties, identifies some of the major problems they entail, and suggests a structural-cognitive approach to urban simulation models.

1 Definitions

The title of this paper refers to two properties that are absent from most AB/CA urban simulation models currently in use: the first is a *structural approach* while the second is a *cognitive approach*. By a structural approach I mean an urban simulation model that explicitly considers the evolving global structure of the city and its role in the dynamics, as revealed, for example, by classical location theory. By a cognitive approach to urban modeling I mean a model that derives its agents' behavior from first principles of human cognition as revealed by cognitive science.

2 Justification

Why do we need a structural approach to urban modeling? I suggest two answers. The first is that we have been carried away too far by the medium we currently employ to model cities, to the extent that (to rephrase McLuhan and Fiore's 1967 book) the medium has become the message. I'm referring to the current popularity of AB/CA models as media to model cities. The advantages of these media are

well recorded. On the other hand, however, the inner logic of these models is such, that their emerging global structure has no role in their dynamics.

Why do we need a cognitive approach to urban dynamics and urban simulation models? Or, put the question differently, why do we need urban simulation models that derive the behavior of their agents from first principles of human cognition as revealed by cognitive science? Because the assumptions from which we derive the behavior of agents in our standard AB/CA models are essentially wrong; because of the finding that agents' behavior in the city is strongly influenced by their 'image of the city' and by the global-structural properties of the city. Let me elaborate.

3 Forgotten Cities

AB/CA urban simulation models that currently dominate the field of city modeling can be seen as a second generation of urban models (Batty, this volume). The first generation was developed during the 1960s and 1970s in the context of location theory. The vast majority of location theory city models were top-down in their structure focusing their attention and interest on the global structure of the city: Dominated and inspired by von Thünen's (1966) type land use theories, Christaller's (1966) and Losch's (1954) central place theories, gravity/spatial interaction models and the like, urban models described cities in terms of center periphery, hierarchies of central places, systems of cities, rank-size distributions and the like (a good representative of this approach is Haggett et al. 1977). Despite the justified criticism that was leveled at these theories by the social theory oriented structuralist-Marxist and humanistic geographies (SMH), and by complexity theory oriented urban studies (Portugali 2006), one cannot deny nor ignore the significant insight about cities and urbanism produced by this first generation of urban theories and modeling. This insight centered as noted on the global structure of the city. There were of course a few exceptions, such as Hagerstrand's (1967) spatial innovation diffusion theory, that were bottom-up in their structure. But they were developed in the top-down context of location theory and thus created a good balance between the two views on the city. In fact, in a more developed version of his theory Hagerstrand showed how the bottom-up spatial diffusion process is influenced by the global structure of the city (by its central place hierarchy).

The story of the second wave of urban models started with the pioneering studies of people like Peter Allen (1981) and Weidlich (1987) who have applied and introduced the theories of complexity to the domain of cities; Allen has applied Prigogine's theory of dissipative structures (Nicolis and Prigogine 1977) while Weidlich Haken's (1983) theory of synergetics. Looked upon in terms of top-down/bottom-up dichotomy, complexity theories can be said to hold the city in its two edges: they are built to show how local processes and interactions give rise to global structures and how the latter feed back and affect local processes, and so on in circular causality. This is the essence, for example, of Allen's urban

simulation model that reformulates central place theory in terms of Prigogine's dissipative structures.

Allen and Weidlich have developed their urban simulation models by means of the mathematical formalisms of dissipative structures and synergetic. The advantage of such a formulation is that it allows capturing in one model both top-down and bottom-up urban dynamics. Their disadvantage is their complicated mathematical structure. 'The use of differential equations in models of self-organization', writes Mikhailov (1990 p. 40),

often makes the computation very tedious ... In this situation ... we can abstain from a numerical integration of exact differential equations and turn to the analysis of much simpler systems represented by networks of cellular automata.

The advantage of AB/CA models is their simplicity, straightforwardness and intuitive clarity. Their disadvantage is that the global parameters of the systems under consideration are left out of the dynamics. In the domain of cities the advantages outweighed the disadvantages not only because of the above mentioned reasons, but also because of two additional ones: First, the cellular structure of AB/CA models and their basic logic correspond nicely our intuition and knowledge about cities: cells represent city lots while the CA dynamics by which the properties of a cell is determined as a function of its nearest neighbors correspond to what is known about relations between elements in the city – the land value of a given lot is strongly influenced by the properties of its nearest neighbors, for instance.

Second, Allen and Weidlich models have attracted mainly students of cities that had an inclination toward the so called 'quantitative approaches'. Despite their inclination, however, most people in this community were and are mathematically non-experts and thus had/have only limited access to the "numerical integration of exact differential equations". AB/CA models with their "much simpler systems represented by networks of cellular automata" enable this community to become full partners in the development of and discourse on urban simulation models.

The result of the above is that the cities of classical location theories with their core periphery and hierarchical central place structures are nowadays forgotten. Forgotten also are the cities of self-organization theories with their dissipative structures and order parameters that emerge out of local interaction but that once emerged enslave their agents. Forgotten also is the fact that AB/CA models were designed to study only half of the process of the self-organization of cities (the process by which local interactions give rise to global patterns), leaving the second half (the process by which the global structure of cities influences agents and local interactions – the synergetics' "enslaving principle") to the more complex and tedious formalisms of Prigogine's dissipative structures and Haken's synergetics.

4 Imagined Cities

We need a cognitive approach to cities, so was claimed above, because the assumptions from which we derive the behavior of agents in our standard AB/CA models are essentially wrong; because agents' behavior in the city is strongly influenced by their 'image of the city' that refers, in fact, to the global-structural properties of the city.

I have discussed these issues in two previous studies: one that suggests a cognitive approach to urban modeling (Portugali 2004) and another suggesting that complexity theory has the potential to bridge between place and space (Portugali, 2006). This section synthesizes these two studies. A good way to start conveying the need for a structural-cognitive approach to urban simulation modeling is Simon's Ant hypothesis.

In his *The Sciences of the Artificial* Simon (1969, 1999) attempts to show how the studies of artifacts can become sciences – the sciences of the artificial. He starts by noting that the key issue that enabled transforming the *study* of nature into a natural *science*, was the finding that a *few* purposeless natural laws that govern the interactions between *few* and *simple* elementary parts, generate the enormous complexity of nature. He then claims that the major obstacle in transforming the study of artifacts into science is our taken for granted twofold assumption that (1) the apparent complexity of human behavior reflects the innate complexity of the human mind, and (2), this innate complexity is the property that makes us 'human' and thus forms the boundary between the 'natural' and the 'artificial'. But this taken for granted assumption simply deceives us, says Simon, because it is only an external appearance of an innately simple behaving system – innately, he claims, we are very simple, not very different from ants. He develops this claim by means of two constructs: the two part "Ant Hypothesis" and the distinction between humans' inner and external environment (I- and E-environments, respectively).

Simon's Ant Hypothesis

Hypothesis I:

"An ant, viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself." (p. 52)

Hypothesis II:

"Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves." (p. 53)

From the Ant Hypothesis follows a distinction between two human environments: an inner (I) environment that is very simple – like the ants, and an externally observed environment that appears to us complex. Based on this distinction, Simon suggests the following structure of explanation for the sciences of the artificial:

I- and E-environments

Humans' Innate and simple I-environment, in the form of intentions (I), aims (A) and needs (N), implemented in, or rationally adapted to, the complex E-environment by means of plans (P), design (D) and engineering (E), gives rise to the observed complex behavior (B) as complex artifacts.

That is,

$$B(\text{observed}) = f(I,A,N,P,D,E \dots)$$

The advantages of Simon's two constructs (the Ant hypothesis and his distinction between I- vs. E-environments) to the dynamic of cities are apparent: They enable a causal-scientific treatment of cities, guided by the Occam's Razor principle according to which 'all things being equal, the simplest explanation is the correct one.'

$$\text{simple cause}(s) \rightarrow \text{complex effect}(s)$$

This is in fact the logic and structure of CA models in general and of the majority of our AC/CA urban simulation models in particular: they start with simple agents that by means of their local interactions give rise to the global pattern of the city. As will be further emphasized below, one of the outcomes of the above is that the global structure of the city is just an outcome; it loses the primacy it has in location theory, for example.

5 The Ghost in the Machine

Philosophically, Simon's thesis conforms to Dualism – to what Ryle (1949) has termed the "ghost in the machine" – the view that human beings are comprised of a tangible body and an intangible mind. Applied to cities and urban simulation models, the dualist position corresponds to the view that agents' rules of behavior (intentions, aims, ...) and the properties of the city are independent from each other; hence the possibility of causal relations between the two. This shows up in the majority of current AB/CA urban simulation models (Fig. 1): they start, for instance, with agents that come to a new city in order to find a location in it. Similarly to ghosts, such agents have no physical properties that are relevant to the urban dynamics – only aims, wants or needs. The agent first considers the state of the city's cells, identifies the empty ones, selects the best of them and locates itself in it. This location act changes the properties of the neighboring cells as is usual in CA models. As illustrated in Fig. 1, from the latter model stage follow two lines of information: one that feeds back to the starting point of the model – to the state of the cells, and another that goes to the researcher and concerns the state of the city as a whole. Note that the state of the global structure of the city has no role in the model's dynamics.

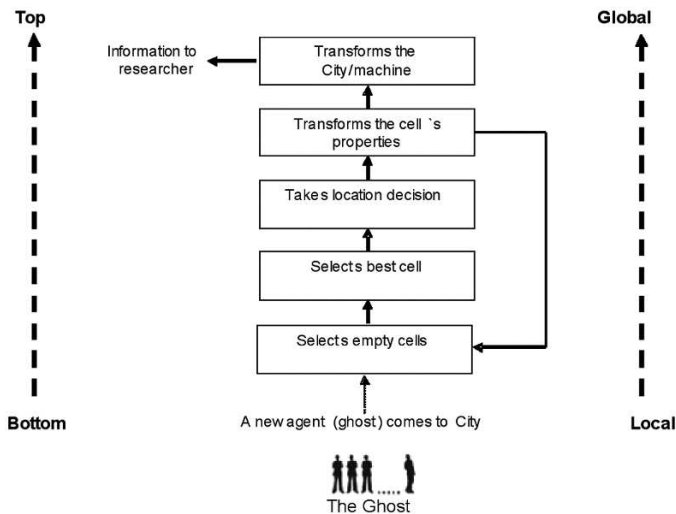


Fig. 1. AB/AC urban simulation models that conform to the ‘ghost in the machine’ doctrine

In *The Concept of Mind* Ryle (1949) argues against the dualist view, suggesting that the mind is not a non-physical substance residing in the body, “a ghost in the machine”, but a set of capacities and abilities belonging to the body. The dogma of the “ghost in the machine” is a derogatory term coined by him to abuse Dualism. Applied to cities and urban simulation models, Ryle’s anti-dualist position corresponds to the view that agents’ rules of behavior and the properties of city within which they act and are situated, cannot be detached from each other; rather, agents’ rules of behavior emerge out of their interaction with the specific properties of the urban environment.

6 Problems

The dogma of ‘the ghost in the machine’, Simon’s Ant hypothesis, and the AB/CA urban simulation models that are build along these lines, are problematic in several respects: They get no support from empirical studies on agents’ behavior, they contradict the paradigm of embodied cognition that dominate other domains were AB/CA simulation models are employed (artificial intelligence – AI, and artificial life – A-life, for instance), and on top of these they contradict the basics of complexity theory and of the city as a complex self-organizing system.

6.1 Problems Due to Empirical Studies on Cognitive Maps

Empirical studies on animals’ and humans’ behavior that are specifically relevant to our understanding of agents’ behavior in the context of urban simulation mod-

els, falsify Simon's Ant hypothesis and by implications its dualist world view. Let us first have a look at Tolman's (1948) notion of *cognitive maps*.

In "Cognitive maps in rats and men" Tolman (1948) summarizes a set of experiments demonstrating that rats and humans have the capability and tendency to construct in their minds cognitive maps, that is to say, information referring to the global structure of the environment. Six decades of cognitive maps studies have indicated that humans conduct their behavior in cities on the basis of their cognitive maps of them. These studies further show that from the notion of cognitive maps follows that urban agents never come to a new city *tabula rasa*. Rather, they come equipped with two types of cognitive maps (Portugali 2004; 2005a): *C-cognitive maps* that are category-like and refer to agents' perception of a *City*, and *s-cognitive maps* that refer to agents' perception of specific cities. Both *c-* and *s-*cognitive maps refer to the global structure of cities with the implication that a substantial part of agents' location decisions, behavior and action in cities is taken in a top-down manner.

Problem 1: Empirical and theoretical studies on cognitive maps indicate that a lot of human behavior in cities is top-down in its structure and yet, the vast majority of AB/CA urban simulation models are bottom-up in their structure.

6.2 Problems Due to Empirical Studies on Exploratory Behavior

The second set of empirical studies concerns the notion of *exploratory behavior*. Experiments in this domain show that animals and humans have an innate drive to explore new environments when first introduced to them. Exploratory behavior studies indicate that exploration is implemented by specific patterns of behavior in space (Golani et al. 1999). These studies further indicate that this innate drive is highly complex. This conclusion follows the finding that when introduced to an environment whose spatial structure is complex, the animal performs a complex exploratory behavior that uses and takes advantage of, the complexity of that environment. But when introduced to a spatially simple environment, the animal still performs a complex behavior that in fact transforms the environment from simple to complex for that specific animal. (For detailed description and bibliography see Portugali 2002).

Problem 2: AB/CA studies are built on the logic of Simon's ant, namely, they postulate that agents' "innate" behavior is essentially simple and complexity is a property of the emergent urban environment. Studies in exploratory behavior indicate that this is not the case, namely, that agents' rules of behavior are from the start complex.

6.3 Problems Due to the Embodied Cognition Ontology

The notions of Embodied and situated cognition that are now prominent in cognitive science, and the notion of Synergetic Inter-Representation Networks (Haken and Portugali 1996; Portugali 2002), falsify dualism and the Ant Hypothesis.

Agents' rules of behavior in multi agents simulation models can be derived in two ways: by postulating a certain behavior and by reference to empirical data. The latter usually come from cognitive science – the discipline that investigates the relations between mind, body and behavior. The first way is common in economics, while the second in AI and A-life. The practice in most AB/CA urban simulation models is to follow the way of economics and to postulate the behavior of urban agents. The consequences are problems 1 and 2 as described above. The suggestion here is to follow the practice in AI and A-life and to look at cognitive science. This is done below. As we'll see immediately, the two main ontologies that dominate cognitive science are the dualist *classical cognitivism* and the non-dualist *embodied cognition*.

In the 1950s and 1960s, the first two decades of the young cognitive science, the dualist classical cognitivism was almost the sole paradigm. The “enemy” was behaviorism that denied the very existence of mind as a researchable-scientific entity and the main challenge of the *Mind's New Science* (Gardner 1987) was to defeat the behaviorist thesis and to prove that the study of mind can become science. Classical cognitivism is still dominant but not as in the past: Since the mid 1970 we see several non-dualist approaches that include PDP (Parallel Distributed Processing) and neo-connectionism (Rumelhart et al. 1986), pragmatist environmental cognition (Freeman 1999), experiential realism (Johnson 1987; Lakoff 1987), situated cognition (Clancey 1997) and embodied cognition (Varela et al. 1994), among others, that are challenging the classical view. These views, commonly termed the embodied cognition ontology, are challenging the classical view not only on philosophical grounds, as in Tyle's thesis, but mainly on empirical grounds. These empirical studies indicate that “the ghost *is* the machine”, that is to say, that perception is bodily action; that agents come to the world already complex; that agents' cognition is situated in the environment within which they act, and that their cognitive system is a network extending beyond their brains/bodies to include the artifacts they produce.

The attraction of the embodied cognition ontology to AI and A-life – the domains engaged in artificial minds, bodies and environments (Franklin 1997) is apparent: building a robot, for instance, or an artificial VR city, literally forces the builder to be aware and sensitive to the intimate relations that exist between the structure of the body/machine and its perceptual-cognitive capacities, or between the specific properties of artificial VR city and agents' perceptual-cognitive-bodily capabilities in it. (On the difference between cognition and action in VR environments versus real environments see Portugali 2005b, 2005c). AB/CA simulation models are common tools in the domain of AI and A-life, it must be added.

Problem 3: Applied to AB/CA urban simulation models embodied cognition as employed in the domain of AI and A-life implies that agents' rules of behavior are

embodied and situated, namely, they are not fixed nor pre-determined. Rather, they are emergent forms of adaptation to specific task and the properties of the environment. The vast majority of AB/CA urban simulation models tend to ignore the above and to postulate agent's behavior rather arbitrarily.

6.4 Problems Due to Complexity Theory

AB/CA simulation models are usually employed as convenient tools to study cities as complex systems. And yet, complexity theory negates dualism: in place of the dualist idea that issues should always be divided into dichotomized states, such as cause/effect, mind/matter, etc., complexity suggests a non-dualist interactive modes of thought that take into account the wider connectivity issues and the need to balance multiple forces.

A case in point is Haken's (1983) theory of *synergetics* that in place of simple cause-effect relations emphasizes complex *circular causality* and in place of mind that gives orders (causes) to the body that then behave in the environment, synergetics suggests a single mind-body-environment system (see the notion of SIRM below); and in place of the Ant's hypothesis' simple cause → complex effect, it suggests a new form of information and complexity reduction.

6.4.1 Self-Organization and the City

Haken synergetics was the main source of inspiration to *Self-Organization and the City* (Portugali 1999). This project attempts to adapt – not just apply – the properties of complexity and self-organization that originated in the domain of the “hard” sciences, to the “soft” domain of cities. A central insight that emerges from this project/adaptation is that cities, like languages, are *dual self-organizing systems*: The city as a whole is a complex self-organizing system, and each of the many agents operating in the city is a complex self-organizing system by itself too. As a consequence, unlike Simon's model in which the interaction between simple innate causes and the environment leads, or gives rise, to a complex system, in the dynamic of cities we find a situation by which the interaction between the many local complex urban agents that operate in the city leads, or gives rise, to the global city as a complex system. From the above follows Problem 4, namely:

Problem 4: AB/CA urban simulation models that were originally designed as means to study the properties of cities as complex systems (characterized as they are by circular causality, dual self-organization and the like), are built as if cities are mechanistic systems characterized by simple causality.

7 The Ghost /s the Machine

The main advantage of the dualist ‘ghost in the machine’ thesis and Simon's Ant hypothesis is that they allow causal relations and as a consequence scientific ap-

proach to the study of cities. The above evidences against dualism and the Ant hypothesis indicate that in the domain of cities one should adopt an alternative view, namely, that ‘the ghost *is* the machine’; that processes of perception (ghost), the behavior and action of the body (agent = machine), and the environment within which actions are situated, form a single complex self-organizing system. In light of this view, can there be a scientific approach that allows complex-complex relationship; can there be a scientific approach to cities?

At the background to this question one has to put also the Structuralist-Marxist, humanistic and more recently postmodernist criticisms against quantitative geography in general and the study of cities in particular (for bibliography and discussion see Portugali 2006). Methodologically speaking, the essence of this criticism is that cities are not natural objects but artifacts and as such do not lend themselves to a scientific treatment based on the natural sciences model. Instead of analysis these critics thus suggest hermeneutics.

The answer to the above question is positive: complexity theory suggests an approach that allows treating scientifically complex-complex relations. This approach that is implicit in several theories of complexity, appears explicitly in Haken’s (2002) theory of synergetics. According to synergetics, self-organization is a property of The Natural -- fluids, the brain, the cognitive system, but also of The Artificial – society, economy, language and the city. In place of the natural–artificial dichotomies, that are often employed as distinctions between “sciences” and “studies” (e.g. cognitive science vs. cognitive studies), synergetics suggests new ones: simple-complex, closed-open, local-global, systems and most importantly: a new form of complexity reduction: by means of the order parameter(s). As an example consider Fig. 2 that describes one of synergetics’ main paradigmatic case studies – the *pattern recognition paradigm* (Haken 1991, 2004):

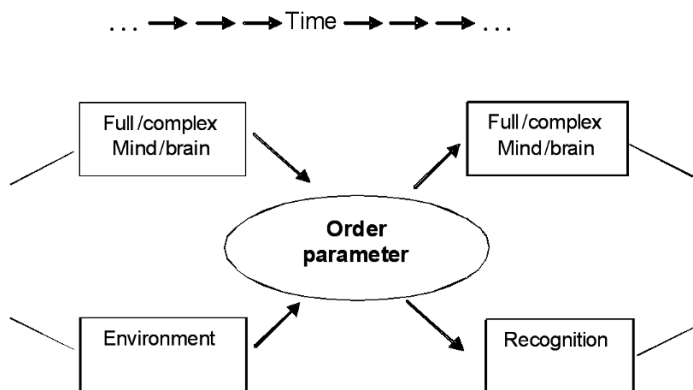


Fig. 2. Synergetics’ pattern recognition paradigm

A typical scenario of the pattern recognition paradigm starts (Fig. 2) when a complex agent with a complex mind full of patterns stored in memory, is offered a

new pattern in a certain environment and is asked to recognize it. According to synergetics, the interaction between the patterns stored in the agent's memory and the offered one in the environment gives rise to an order parameter that enslaves the many parts of which the system is composed. When this is done, recognition is accomplished. As can be seen, the interaction between the many parts of the system gives rise to an order parameter, but once the order parameter comes into being, it prescribes and describes (for the external observer) the behavior of the parts. The result is circular causality and a situation by which the global order of the system – the outcome of the self-organization process – is relatively simple and as a consequence lent itself to interpretation. In other words, instead of the mechanistic process of

simple cause → *complex effect*,

we get

complex “causes” → *simple effect*.

8 SIRN – Synergetic Inter-Representation Networks

The notion of SIRN – Synergetic Inter-Representation Networks (Haken and Portugali 1996; Portugali 2002) is an attempt to develop a theory about agents' cognition, behavior and action that takes into account the fact that a major part of human cognition, behavior and action, concerns the production of artifacts. Such artifact might be bodily and individual as in the case of dance, or bodily and collective as in the case of spoken languages; they also might be stand-alone and the product of an individual person, as in the case of a chair produced by a carpenter, and they might be stand-alone and the product of collective interactions, as in the case of cities – cities are huge, stand-alone artifacts and the product of interaction between a very large number of people, firms, institutions etc. (or 'agents' as we term them in our models). According to SIRN the process of the production of artifacts is an integrative element in the process of cognition. More specifically, SIRN suggests that many cognitive processes evolve as an interaction between internal representations (percepts, images) constructed in the brain and external representations (artifacts) constructed in the environment. Some of the artifacts are small like rings or chairs, while others are large like cities.

A nice way to convey the notion of SIRN and its role in the dynamic of cities is provided by the city game experiment (Fig. 3). The experiment involves a group of some 40 to 70 participants playing a game the aim of which is to build a city on a floor representing the site for a new city. The players sit in a circle around the floor (the “building site”); each of them is given a 1:100 mock-up of a building and is asked to locate it in the virtual city on the floor, in what s/he considers as the best location for that building. The process is sequential in the sense that each of the players locates his/her building in his/her turn. The process is also public in the sense that each agent observes the city as it develops and in the process also learns the spontaneously emerging order on the ground. It is typical in such games

that, after a few initial iterations, an observable urban order emerges. Once this happens, the participants internalize this emerging order and tend to locate their buildings in line with it (Portugali 1996).



Fig. 3. The city game experiment

The main features of such a game are the following:

- A sequential interplay between internal representations constructed in the mind of each of the participants and the external representation, that is the city, constructed on the floor.
- The above interplay illustrates the interaction between what has been defined above as c- and s-cognitive maps: Each player come to the game with a conception of 'a city' that is, c-cognitive map. The interaction between his/her c-cognitive map and the actual city on the ground gives rise to the player's s-cognitive map according to which s/he then takes location decision.
- The game illustrates the emergence of a collective artifact – the city on the floor – as a consequence of the dynamics of a *dual* self-organizing system: Each of the players is a complex-local self-organizing system and the emerging city on the floor is a global complex system.
- The global structure of the artifact city (on the floor) plays the role of the order parameter: it emerges out of the interaction between the individual location decisions of the players. But once it emerges, that is, once an identifiable global city structure could be observed on the floor, that order enslaved (in the sense of synergetics) the subsequent decisions and actions of the participants.

This latter point takes us full circle to the port of departure of this paper, namely, to the section “forgotten cities” that claims that current AB/CA urban simulation models tend to overlook the global structure of the city. What comes out from the discussion here is that overlooking the role of the global structure is wrong from the point of view of human/agents’ behavior in the context of cities. As was noted above, Allen and Weidlich paid full attention to, and theorized about, the global structure of the city. But they did so at the expense of local-cognitive dynamics. Both Allen and Weidlich postulate the behavior of agents; Allen from the perspective of dissipative structures and Weidlich by means of the master equation approach. An attempt to take both aspects into consideration is Haken and Portugali’s city model that approaches the dynamics of cities from the perspective of synergetics’ pattern recognition paradigm (Haken and Portugali 1996; Portugali 1999 Chap. 13). The question that is relevant to the present discussion is this: ‘in light of usefulness, popularity and prominence of AB/CA urban simulation models, can there be an approach that will add to bottom-up AB/CA urban simulation models a top-down componet? The model CogCity that is introduced next answers this question in the affirmative.

9 Cognitive City

Cognitive city or in short *CogCity*, is a structurally and cognitively oriented AB urban simulation model. In is cognitive in the sense that it derives its agents’ behavior from first principles of human cognition as revealed by cognitive science, and it is structural in the sense that it explicitly considers the evolving global structure of the city and its role in the dynamics.

Agents’ behavior in CogCity is in line with SIRM and the city game as described above. Fig. 4 illustrates the process in relation to synergetics’ pattern recognition paradigm, while Fig. 5, in relation to what one might call the basic scenario of the CogCity model.

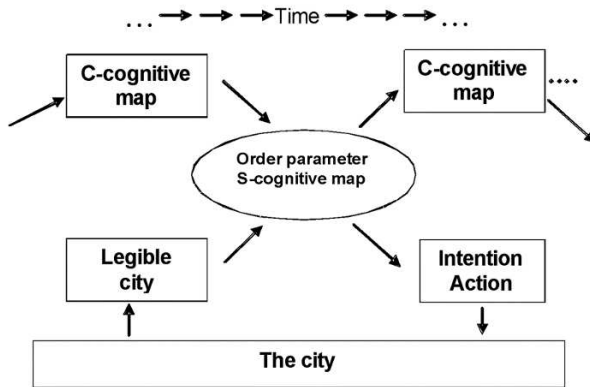


Fig. 4. Agents’ behavior in CogCity is in line with SIRM

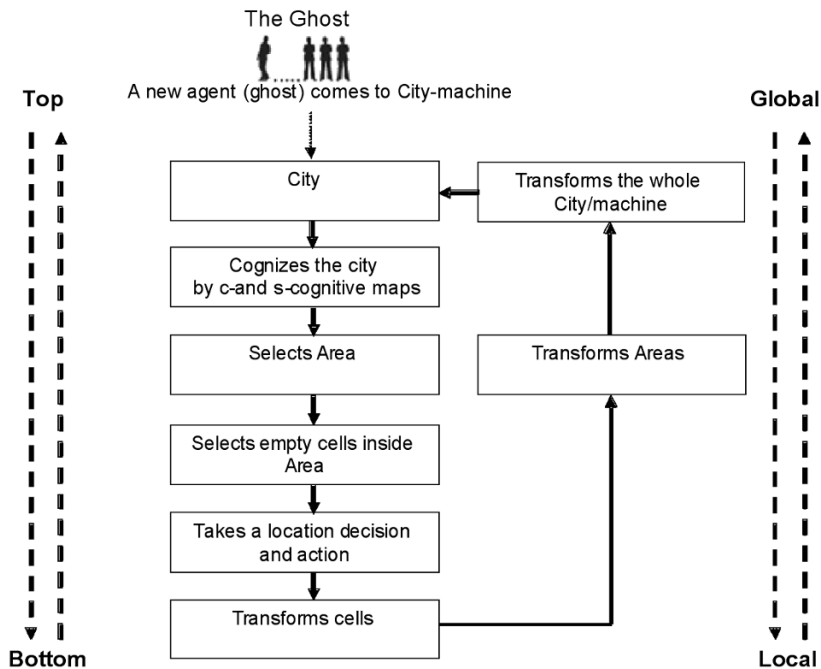


Fig. 5. The basic scenario of CogCity

The basic scenario or model flow of CogCity can be described as follow:

- Agents come to a new city in order to find a location and live in it. Each agent has some image of a city in its memory, that is to say, a c-cognitive map that reflects the agent's past experience and knowledge of cities.
- Each agent observes, pattern recognizes and interpret, the city by means of its c-cognitive map – by comparing its c-cognitive map of 'a city' with the city on the ground. The result is the agent's s-cognitive map.
- By means of its s-cognitive maps the agent takes location decision hierarchically: it first decides the area in the city proper for its task (to build a residential building, factory, office building, etc.); it then considers the empty cells in that area; then the properties of the cell (its neighbors, etc.).
- Each agent then takes decision and action locally
- The synergistic interaction between the many agents gives rise to the global form and structure of the city
- The latter feeds back on the cognitive maps of individual agents
- And so on in circular causality

As was noted above, in standard AB/CA urban simulation models the emerging global structure of the city is an outcome – it doesn't feed back (or forward) to the

dynamics of the city. In the above scenario, per contra, it is central to the dynamics: it plays a role in shaping the s-cognitive maps of agents – the information according to which they take location decisions. In other words, CogCity requires from the modeler to explicitly consider the city infrastructure. CogCity employs for this purpose some of the good-old-classical notions, that as a consequence of the dominance of the new media – AB/CA (and the quantitative revolution in geography: humanistic, structuralist-Marxist, postmodernist) – were almost forgotten: centers and central places, periphery, the notions of range, areas and so on. CogCity considers the modeled city as a hierarchical structure composed of *Areas* (and *Subareas*), built of *Central Places* and their *Peripheries* defined by the central place's *Range* that measures the central place's intensity. It then employs the *rank-size* distribution (“rule”) as means to define the global structure of the city.

10 Concluding Notes

There is no room here to further elaborate the detailed properties of CogCity – the interested reader can find such details in Portugali 2004. What is important to note here, however, is that while AB/CA models were originally designed to model bottom-up process and to leave the more complex top-down dynamics to mathematically more complicated models, they can still be constructed to take top-down process into consideration. The model CogCity demonstrates that a AB/CA urban simulation model that combines top-down and bottom-up processes in one model is possible and useful.

References

- Allen P (1981) The evolutionary paradigm of dissipative structures. In: Jantsch E (ed) *The Evolutionary Vision*. Westview Press, Boulder, pp 25-71
- Clancey WJ (1997) *Situated cognition*. Cambridge University Press, New York
- Christaller W (1966) *Central Places in Southern Germany*. Prentice-Hall, Englewood Cliffs, NJ
- Franklin S (1997) *Artificial Minds*. MIT Press, Cambridge, MA
- Freeman WJ (1999) *How Brains Make Up Their Minds*. Weidenfeld & Nicolson, London
- Gardner H (1987) *The Mind's New Science*. Basic Books, New York
- Golani I, Kafkafi N, Drai D (1999) Phenotyping stereotypic behaviour: collective variables, range of variation and predictability. *Applied Animal Behaviour Science* **65**: 191-220
- Hagerstrand T (1967) *Innovation Diffusion as a Spatial Process*. University of Chicago Press, Chicago
- Haggett P, Cliff AD, Frey A (1977) *Locational Models*. Edward Arnold, London
- Haken H (1983) *Synergetics: An Introduction*. Springer, Heidelberg
- Haken H (1988, 2000) *Information and Self-Organization: A Macroscopic Approach to Complex Systems*. Springer, Heidelberg
- Haken H (1991, 2004) *Synergetic Computers and Cognition*. Springer, Heidelberg

- Haken H (1996) *Principles of Brain Functioning*. Springer, Heidelberg
- Haken H (2002) Complexity reduction. 10 Herbstakademie conference on Self-Organization, Kloster, Seeon.
- Haken H (1996) *Principles of Brain Functioning*. Springer, Heidelberg
- Haken H, Portugali J (1995) A synergetic approach to the self-organization of cities. *Environment and Planning B: Planning and Design* **22**: 35-46
- Haken H, Portugali J (1996) Synergetics, Inter-representation networks and cognitive maps. In Portugali J (ed.) *The construction of cognitive maps*. Kluwer academic publishers, Dordrecht, pp. 45-67
- Johnson M (1987) *The Body in the Mind: the bodily basis of meaning, imagination, and reason*. The University of Chicago press, Chicago
- Lakoff G (1987) *Women, Fire and Dangerous Things: what can categories reveal about the mind*. The University of Chicago press, Chicago
- Losch A (1954) *The Economics of Location*. Yale University Press, New Haven, CN
- McLuhan M, Fiore Q (1967) *The Medium is the Massage*. Penguin books, London
- Mikhailov AS (1990) *Foundation of Synergetics I*. Springer, Heidelberg
- Nicolis G, Prigogine I (1977) *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations*. Wiley, New York
- Portugali J (1999) *Self-Organization and the City*. Springer, Heidelberg
- Portugali J (2004) Toward a cognitive approach to urban dynamics. *Environment and Planning B, Planning and Design* **31**: 589-613
- Portugali J (2006) Complexity theory as a link between Space and Place. *Environment and Planning A* **38**: 4: 647-664
- Portugali J (2002) The seven basic propositions of SIRN (Synergetic Inter-Representation Networks). *Nonlinear Phenomena in Complex Systems* **5**: 4: 428-444
- Portugali J (2005a) Cognitive maps are over 60. In: Mark D, Cohn A (eds) *Spatial Information Theory, COSIT 2005*. Springer Lectures in Computer Science series, Berlin
- Portugali J (2005b) The scope of complex artificial environments. In: Portugali J (ed) *Complex Artificial Environments*. Springer, Complexity series, Heidelberg
- Portugali J (2005c) Complex artificial environments – the ESLab’s experience. In Portugali J (ed) *Complex Artificial Environments*. Springer, Complexity series, Heidelberg
- Portugali J (1996) Inter-representation networks and cognitive maps. In: Portugali J (ed) *The construction of cognitive maps*. Kluwer academic publishers, Dordrecht, pp 11-43
- Rumelhart DE, Smolensky P, McClelland JL, Hinton GE (1986) Schemata and sequential thought processes in PDP models. In: McClelland JL, Rumelhart DE, PDP Research Group (eds) *Parallel Distributed Processing, Explorations in the Microstructure of Cognition*. vol 2: Psychological and Biological Models. MIT Press, Cambridge, MA, pp 7-57
- Ryle G (1949) *The Concept of Mind*. Hutchinson’s University Library, London
- Simon HA (1981/1999) *The Sciences of the Artificial*. MIT Press, Cambridge, CA
- Tolman E (1948) Cognitive maps in rats and men. *Psychological Review* **56**:144-155
- Thunen JH von (1826, 1966) *The Isolated State*. Pergamon, Oxford
- Varela FJ, Thompson E, Rosch E (1994) *The Embodied Mind*. MIT Press, Cambridge MA
- Weidlich W (1987) Synergetics and social science. In: Graham R, Wunderlin A (eds) *Lasers and Synergetics*. Springer, Berlin, pp 238-256

The Socio-Spatial Dynamics of Systems of Cities and Innovation Processes: a Multi-Level Model

Denise Pumain

Unité Mixed de Recherche Géographie-cités, University of Paris I, France

Abstract

Following a first attempt presented as the SIMPOP model (a multi-agents systems whose prototype is described in Bura et al. 1996), our aim is to develop a generic model for simulating the evolution of systems of towns and cities, using the SWARM simulation platform. The scientific issue is: to understand how cities that are interconnected through material and immaterial networks co-evolve, within an environment where social and economic innovation continuously emerge, while maintaining at a macro-geographical scale functional, hierarchic and spatial differentiation which evolve at a much slower pace.

The SIMPOP2 model is designed for testing hypothesis about the general processes of urbanisation and interactions between towns and cities. The objective is to identify and order the rules and parameters that have produced a variety of configuration at the level of the systems of cities, according mainly to the changing conditions of spatial interaction: communication means, transportation speed, range of trading activities, proximity networks and long distance connectivity. Three main varieties of urban systems that have had different histories of urbanisation and conditions of circulation will be investigated: developed countries with old settlement systems, developed countries of much more recent urbanisation, and developing countries. A first generic version of the model includes the minimal rules that seem necessary for reproducing the emergence and evolution of any system of cities, whereas three different scenarios will be constructed for simulating the characteristic features of the three main variations.

1 Introduction¹

Major modelling efforts in most disciplines of urban sciences have been directed towards understanding and monitoring problems related to one city only. Of course many problems of the same nature, as growth management, economic de-

¹ The SIMPOP2 model is developed in the framework of the European ISCOM programme, with the participation of Anne Bretagnolle, Cécile Buxeda, Clarisse Didelon, Alexis Drogoul, Jean-Marc Favaro, Benoît Glisse, Hélène Mathian, Fabien Paulus, Lena Sanders and Céline Vacchiani.

velopment, land use planning or social cohesion issues can be encountered in many different cities and receive similar solutions, modulated according to various local differences, either quantitative (size) or qualitative (culture). However, modelling urban systems at the city level only is not enough for capturing the essential part of a city's dynamics. We consider cities as "systems within systems of cities" (Berry 1964), because geographers demonstrated since a long time that the development of any city could not be separated from its interactions with other urban entities. Even if cities as complex systems can be modelled as emerging from interactions between multiple decisions taken by institutions and individuals, their development, especially when the long term is considered, is intrinsically constrained by their interdependencies with other towns and cities. These interactions between urban entities are of a different nature from the "external causes" which are usually considered as a backdrop of external conditions in a model, because the interurban interactions directly interfere with the dynamics of the city itself, and are defined according to the city's relative situation in a system of cities.

This multilevel approach seems to be a more specific contribution from geography to urban modelling. We will recall first which main theoretical results (actually, generalisation from many comparative empirical observations across space and time) can be integrated in a model, before presenting our choices in designing a multi-level version of a multi-agent system model named SIMPOP2.

2 An Evolutionary Theory of Systems of Cities

Urban systems are social systems which kept remarkably pervasive features over historical times: an elementary observation is that we use since more than two thousand years the same words of towns or cities for designating permanent agglomerations of people and settlements whose form, size, functional content and social meaning have been considerably transformed, especially during the last two centuries. I have suggested elsewhere to consider urban systems as territorial adaptors to social change (Pumain 2000), and demonstrated that the structure and the dynamics of urban systems could not be separated from the historical common trends which govern their evolution. Especially, the expansive trend which characterize these systems, not only in a quantitative way (growth in population and wealth) but also qualitatively (through technological and social innovation) is an essential driving force of their structure and evolution.

2.1 The Urban Transition as an Emergence Process

W. Zelinski (1971) suggested to call "urban transition", an expression analogous to "demographic transition", that can be interpreted as a phase transition in physics, the transformation of the world pattern of settlement, from a mainly rural habitat, constituted of relatively homogeneous and dispersed small population clusters, into an urban form of habitat, made of much larger, concentrated and dif-

ferentiated units. According to the region of the world, the transition occurred more or less early in time (from beginning of 19th century until about 1950) and with different paces. But everywhere it was accompanied by an intensification of the communication between towns and cities, by multiplying the networks that connect them, especially through the gains in speed provided by new means of transportation and communication. Meanwhile, the transition everywhere lead to huge increases in city sizes (more than 30 urban agglomerations are today above 10 millions of inhabitants) and to a wider concentration of population within cities (half of the world population is urban in 2000, 80% in developed countries). Moreover, the evolution of these towns and cities is so coherent that, first within regional then national frames (and even later at world scale), is has been suggested to consider them as forming *systems of cities*, as early as in 1841 (Jean Reynaud quoted by Robic 1982). The emerging properties of such systems can be interpreted as resulting from the interactions between cities (Pumain 1997).

Studying the properties of such systems may bring an important contribution to the issue of the future of urbanisation: on the long run. One may wonder if, after the end of the urban transition, settlements will continue to concentrate and to differentiate, or if a bifurcation toward a different form of organisation will occur. On a shorter period of time, a better knowledge of the dynamics of urban system can help to predict the effects of the increasing connections linking towns and cities, either from political decisions (as following the European integration process) or, more generally, due to the diversity of linkages created through the globalisation process. The main question then is about how cities are going to redefine their relative positions within enlarged networks and hierarchies (Taylor 1997). On the very short term, many questions remain debated, as the possible effects of the new means of communication and trade on urban systems: are the NICT going to overturn the existing urban hierarchies, or will they be progressively integrated into these systems without changing much their relative configuration, as this happened in the past for previous innovations (telegraph, railway, telephone, automobile...). A generic model that would succeed in simulating on the long term the transformation of urban systems, respective to their capacity of reaction and adaptation to the changing conditions that allow for and constrain communication and trade, would be a great help in understanding the past co-evolution of cities and predicting the future.

2.2 Urban Systems as Complex Systems

An approach in terms of complex system can improve our understanding of the evolution of systems of cities. We think that some of their properties, which were observed for a long time, even if not yet formalised in such terms, can be interpreted within this framework. The main distinctive feature of complex systems is their ability to exhibit emerging properties, or, as quoted by Batty and Torrens (2001), to give rise to a “surprise” for the observer. But of course there are many possible definitions of complexity, from the realistic (“complexity is an order whose code is unknown”, Atlan 1979) to the constructivist (“complexity is the

number of non equivalent interpretations that can be made by an observer of a system”, Livet 1983). During the last thirty years, new developments in the theory of self-organisation in physics (Prigogine 1996; Haken 1977), evolution of living species in biology and adaptive cognitive systems in economy or social networks (Arthur 1994; Lesourne and Orléan 1998; Anderson et al.1988; Arthur et al. 1997; Weidlich 2000) have changed our representations of system dynamics, especially by emphasizing the conditions of emergence of new structures from local interactions between adapting individuals. The main epistemological questions have therefore shifted from the autonomy of systems relative to their environment, towards the identification of attractors governing their dynamics, and lastly to their capacity of innovation within a context of uncertain and changing rules of social interactions (Pumain 2003).

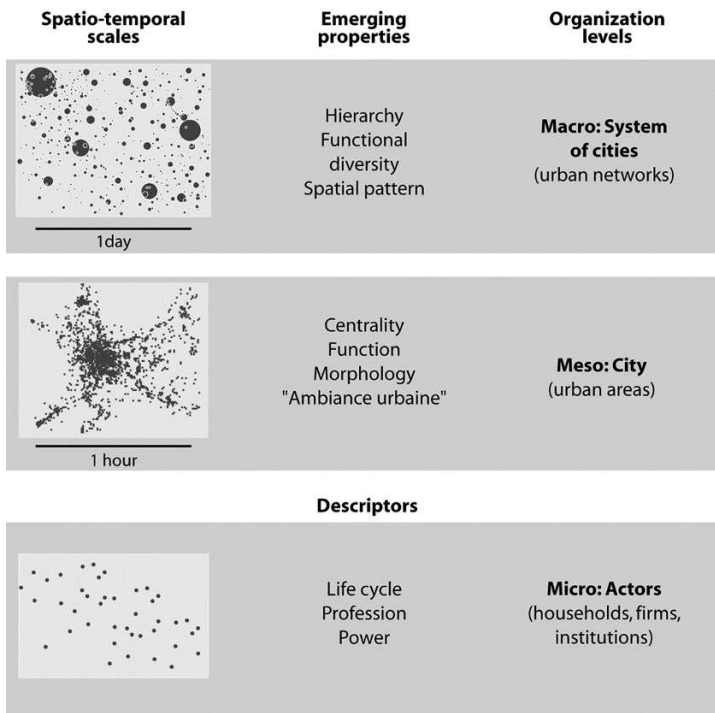


Fig. 1. Urban scales, emerging properties and organization level

Systems of cities can be meaningfully described at least at two levels of observation, that are mainly produced by the interactions occurring at the lower level. Each of them is characterised by emerging properties: indeed *one city* can be described as an organisation that is produced by the interactions between different urban actors, whereas at the upper level *systems of cities* are shaped by the exchanges of persons, goods, capital and information between towns and cities (Figs. 1 and 2). These networks of cities have emerging properties that remain relatively

stable over time and space. A first one is their rather regular spatial organisation (summarized by central place theory), a second related one is the systematic hierarchical differentiation of their sizes (as predicted by Zipf's law), and a third property is their functional differentiation (through economic specialisation, for instance in manufacturing, administration, tourism...). The relative ranking of cities in the system they form (urban hierarchy), as well as their specific economic profile (functional specialisation), can persist over much longer periods of time than the corresponding characteristics of the individuals that are composing them (change in profession, residential migration, replacement of generation in the case of persons or households, creation, closing or merging in the case of firms, renewal of urban institutions...).

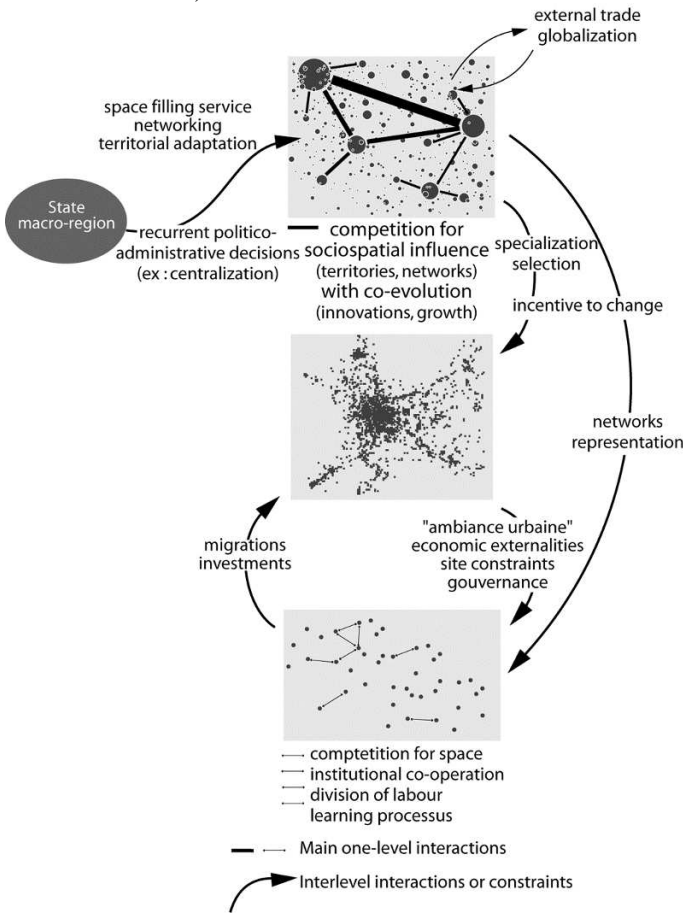


Fig. 2. Interactions between different urban scales

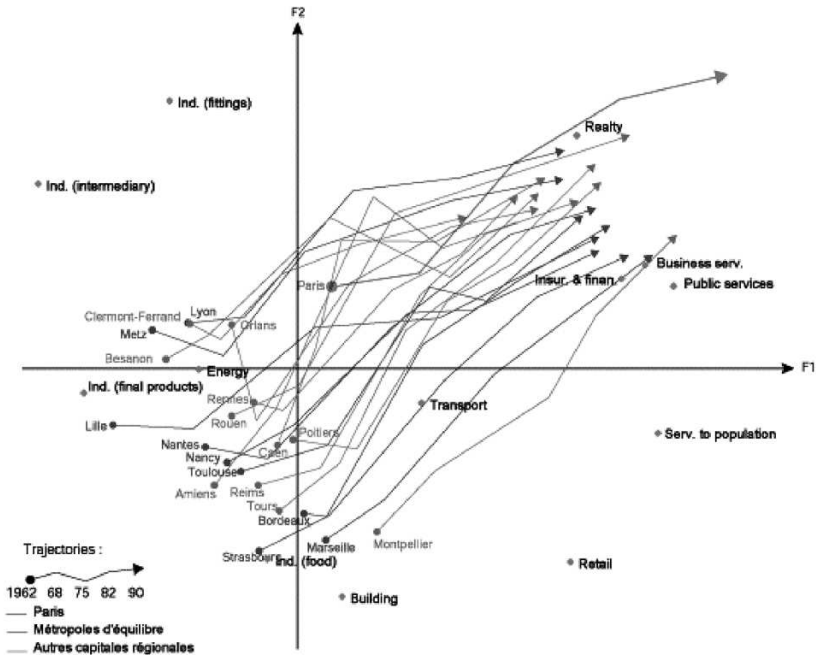


Fig. 3. Trajectories of cities in an economic space. (Source: Fabien Paulus, 2004, *Co-évolution dans les systèmes de villes : croissance et spécialisation des aires urbaines françaises de 1950 à 2000*. Université Paris I, thèse de doctorat)

The evolutionary modes of such systems are similar to those observed for other types of complex systems. The competition between cities lead to many local fluctuations (growth, decline) that in general do not modify the global structure of the system, which remains much more stable (Fig. 3). However that structure is progressively transformed over time, especially through a process of hierarchical diffusion of innovations (concerning for instance in the last two decades the business services, or the cultural industries, which are still concentrated in the largest metropolitan areas), and through cycles of functional specialisation (examples of long distance maritime trade, successive industrial revolutions ...), which gave rise to specific types of cities (Fig. 4). Such processes are essentially non linear: urban growth induces a concentration in the largest cities, partly due to positive feedbacks between size and accessibility (early adoption of modern rapid transportation and communication means, as paved roads, then rapid trains, airlines ...) and partly to the capture of the benefits of the innovations (initial advantage, associated to a site or a situation). During the course of time, the intermediary steps are short-circuited, because of the more limited number of necessary stops within the rapid networks that hampers the chances of development of smaller towns and systematically weakens them. Moreover, the urban dynamics, when expressed in terms of demographic or economic growth, are likely to show reversals: unlike in

the economic product cycle, there is never a total substitutability in the case of cities, but on the contrary a possible reuse of old urban locations, which may have been momentarily abandoned but become attractive again in the course of a new economic cycle. A well known example among many others is the city of Montpellier, in the south of France, that was a brilliant university town in the 18th century, but completely ignored by the industrial revolution of the 19th, and in spectacular revive after 1950.

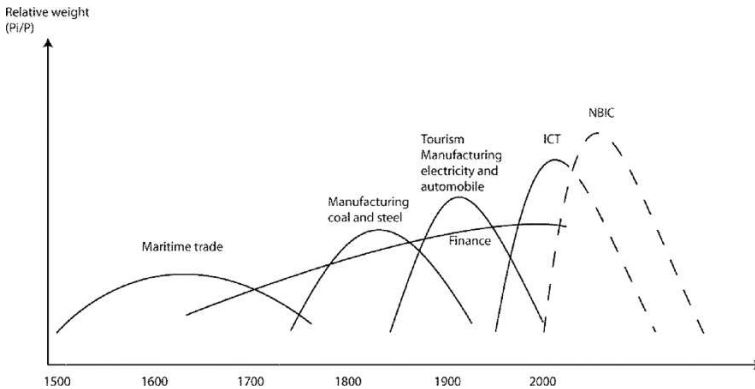


Fig. 4. Main innovation cycles and estimation of relative weight of the corresponding urban specialisation

2.3 Main Steps in Modelling the Evolution of Systems of Cities

Formalising such evolution by dynamic models that enable the emergence of new properties or restructuring of systems has followed a variety of paths. We shall recall only a few steps in an already long history of modelling (Pumain 1998). A precursor can be seen in the first Monte Carlo simulation of the urban growth from rural and interurban migration flows (method Morrill). After a few attempts at using the formalism of catastrophe theory (Casti and Swain 1975; Wilson 1980), a first series of dynamic models expressed as systems of non linear differential equations followed the objective of modelling regional urban systems (Allen and Sanglier 1979; White 1978). These first models described the evolution of state variables at a macro-level, the lower level interactions being summarised in mathematical relationships or in parameters. As interactions are non linear, the systems are not attracted towards a pre-determined equilibrium, the model is used for exploring a variety of possible futures. A shock linked with the amplification of some internal fluctuation, or with an external perturbation, that is, a small change in the parameters of the model, can modify the dynamic trajectory of the systems and persists as a determinant of their further qualitative structure, according to a bifurcation. For instance, a small change in preference of consumers for large size and diversity of shops, as well as a variation in the price of transport,

can produce a spatial concentration of trade in a major urban centre or its dispersion in a multitude of small centres (Wilson 1980; Allen and Sanglier 1979).

Even if some models made analytically more explicit connections between the individual behaviour and the resulting aggregated interactions (as for instance the synergetic model of interregional or interurban migrations first developed by W. Weidlich and G. Haag (1988) using as a starting point a description by a master equation), in practice there was very limited correspondence established with observations at a micro-level, since an “average” behaviour was supposed to be representative of the individuals and the applications were conducted with statistics on aggregated flows (for instance, migrations between cities by Sanders 1992). On the contrary, models of micro-simulation integrated a lot of details about the behaviour and familial or professional career of individuals, but did not pay so much attention to the evolution of the resulting structures at the macro level (Holm and Sanders 2001; Clarke 1996).

Compared to these earlier representations of self-organisation in models, the actual notion of emergent properties refers to a more explicit modelling of individual behaviour and interactions, usually in agent based models or in multi-agent systems. Emergent properties appear at a higher level of aggregation than the original description of the system. They consist most of times in the formation of a structure at the macro level (which may arise in an unexpected way from the individuals and their interactions as they are described in the model), and more rarely consist in the appearance of a true novelty, which was not expected at all, but can be confirmed afterwards in empirical testing. Ideally, this last case would be the prediction of a social innovation. Many applications of complex systems modelling to social sciences are based on interactive cognitive and/or adaptive agents (Bura et al. 1996; Allen 1997).

2 Multi-Agents Systems as Flexible Modelling Tools for Spatial Interaction

Several computing tools have been developed in the last decades for the simulation of emerging properties in systems of interacting agents. Perhaps because the observable species puzzle biologists wondering why imaginable creatures do not exist (Lewontin 2003) imagining artificial life exploring different possibilities is an interesting challenge. From the simplest “life game” conceived by G. Conway (1982) towards the more sophisticated models replicating elements of evolution theory (Wolfram 1994; Langton 1988), a variety of simulations have explored different ways of representing existing or imaginary life processes and their corresponding outcomes. In a parallel way, the computational models of “distributed artificial intelligence” have elaborated task solving methodologies involving different degrees of co-operation between individual agents (Huberman 1988). From the eighties on, they have contributed to develop and enrich concepts as agents, their environment and interactions, receiving new definitions from a variety of applications, in disciplinary fields as economy, ecology, ethology or geography

(Tsfatsion 2002). The “multi-agents systems” rely upon specific software inspired by the object-oriented approach which includes passive attributes and active rules for designated classes of objects; objects belonging to the same class share common processes but may have different data; classes of objects are organised in a hierarchic way according to an heritage principle of their properties.

Multi-agents systems (MAS) are especially useful as simulation tools for modelling a dynamics, when it is essentially explained by the heterogeneity of individual features and their interaction. They enable the modeller to associate qualitative and quantitative rules, and to integrate several levels of organisation as well as diverse time scales and dynamics relationships. They are used in natural, social or cognitive sciences, as artificial laboratories (*in silico*) for observing the behaviour of agents at an individual or collective level, and analysing the evolution of the structures that emerge at a macro-level (Ferber 1995). They appear as a reasonably promising technique for simulating geographic worlds, mainly because of their ability to consider the environment of a system, their acceptance of a wide conceptual diversity of agents (allowing for multi-level analysis) and their flexibility regarding interaction rules, especially in spatial relationships.

Multi-agents systems are much more flexible than differential equations for simulating spatial and evolving interactions, including quantitative and qualitative effects. Through the definition of rules at individual level, they can reproduce the circulation of information between cognitive and decision making agents. They simulate at the upper level the emergence of collective or aggregated structures which can be tested statistically. The rules can be adapted for varying space and time scales of interaction under the course of history. Of course multi-agents systems do not solve the problem of choosing a “good” theoretical representation. A first application to geography was the SIMPOP model (Bura et al. 1996; Sanders et al. 1997), which used MAS for simulating the emergence of a functionally differentiated system of towns and cities from an initial more or less homogeneous rural settlement system, over a duration of some 2000 years. This model differs from most of MAS in two aspects: its agents are immobile, as they represent places in geographical space, even if their interactions are affected by the technical innovations in the communication systems and the increasing level of urban resources which modify the relative situation of elementary cities in the network of their relational space. Second, as SIMPOP is a model of interactions between territories (aggregated geographical objects), the “behaviour” of such agents is not reducible to the behaviour of individual persons, for instance as in models of cognitive economy.

A few other models of differentiated urban growth in a system of cities have been developed recently, especially in economics (within the framework of “new economic geography”), but usually they do not succeed in simulating the emergence of an urban hierarchy (for a thorough discussion see Schweitzer 1998). Different suggestions came from econophysics. For instance, S.E. Page (1998) proposed an agent-based model to simulate the emergence of cities, from very simple assumptions about the location behaviour of agents, conditioned by a preference for agglomeration and an average distance in relation to other agents. However, in this model an agent’s utility is defined, in reference to the distribution of agents on

the lattice, and it is not very plausible that real agents could possess this sort of information. Axtell and Florida (2000) provide a more detailed microeconomic multi-agent model of endogenous formation of business firms, allowing agents to move between firms and between clusters of firms (thus assimilated to cities). Under the hypothesis of increasing returns from clustering at the level of the firm, they simulate a size distribution with constant returns (average growth rate) at the aggregate level. A stationary macro-structure is generated from a non-equilibrium microeconomic process. However appealing, because it reconciles two apparently contradictory but observed processes (search for increasing returns at the individual level, no decisive increasing returns at the aggregate level), this model has not been validated from empirical observations.

Andersson et al. (2003) use an algorithm generating «scale-free» networks. This corresponds to a class of growing networks whose node degrees are power-law distributed (Barabási 2002). In their model, the nodes of the network represent pieces of land which over time become more and more connected by edges representing exchanges of goods and services (the result of this trade is in fact simulated by a trade benefit or financial investment directed from one node to another). The model proceeds by adding new links between already developed nodes, with a probability of this occurring that is proportional to the relative size of the node in the total number of nodes, and by selecting new nodes. The mean probability of developing existing nodes is significantly higher than that relating to the development of new nodes. Spatial rules are added to specify this selection process, according to hypotheses about a distance-decay interaction model. Thus it is not quite clear whether this model is designed to simulate the urbanisation process on the intra-urban scale, or the formation of urban hierarchies, or both (Pumain 2004). In any event, the concept of “scale-free” networks, or “small worlds”, seems well adapted to the simulation of urban systems, since they reveal the hierarchical structure that emerges in progressively constructed networks.

3 SIMPOP2 as a Multi-Level Model

SIMPOP is designed for simulating the emergence, structuring and evolution of a system of cities, starting from an initial spatial distribution of settlements in a large region, state or set of states, a repartition of resources which can be assessed randomly or exogenously, and rules defining how cities interact, grow and specialise. In this model the environment is represented by cells having different kinds of resources for agricultural production or industrial purposes (exploited only from year 1800 on), and with various facilities or obstacles for circulation. They can be allocated randomly or according to a specific pattern. Towns are emerging as centres of accumulation of population and wealth, through first the trading of agricultural surplus in their surrounding cells, then from their competition for the acquisition of other urban functions, as other types of trades, or administrative roles. Interurban competition is simulated by relating profits (from trade or taxes) and by growth rates with a random factor. Meanwhile, the spatial

range of interactions is increased when cities acquire new functions and with technological progress going. As a result, different patterns of towns and cities in terms of spatial and hierarchical distribution are emerging.

The second version of the SIMPOP model, named SIMPOP2 is now experimented. It is now conceived for a larger number of agents (from a maximum of 400 to several thousands). It also includes a better representation of the competitive interaction between cities, through the introduction of two new agents which make explicit the role of the functions of innovation and governance within the dynamics of the urban system. This is a way to complete the theory of urban systems within the framework of complex systems theory, by substantiating the growth process in social terms. What make cities growth? Since Schumpeter, innovations, especially entrepreneurial ones, are theorised as making the economic basis for further urban growth. But if the process of diffusing innovation is well documented, since mainly Hågerstrand's work (1952), the question of its appearance remains much more difficult. A recent review of the literature on industrial districts, innovation and learning processes in regional and urban systems (McKinnon et al. 2003) underlines many uncertainties in our actual knowledge about such processes. The SIMPOP2 model may help in establishing a few dynamic conditions of this emergence of "second nature" advantages.

3.1 Interurban Competition and Innovation as the Main Driving Forces

The SIMPOP model is evolutionary because a major underlying hypothesis is that the pervasive structural features of urban systems that we can observe are produced by an historical evolution. This evolution involves systematic, time-oriented changes in major circumstances of the system over time, including the demographic and urban transitions, the increase in gross and per capita economic wealth, the trendy increase in the speed of transportation means, as well as the recurrent appearance of technical, economic and cultural innovation. Thus, it is this social, historical evolution that supports the dynamics of urban systems, even if in a concrete way it is made through the mechanism of interurban interactions. Those are the "bottom-up" processes leading to the emergence of the structure of the system, whereas the evolutionary trends can be thought of as emerging trends, which are produced as feedback effects by the system of cities itself, and become new constraints on the dynamics of individual cities. Actually, we do not know yet how to make these large evolutionary trends to emerge, and they are represented in an exogenous way within the model, whereas it is possible to represent the endogenous process of building an urban hierarchy from the interactions between cities.

Interaction between cities keep over time some permanent features, among them the most important is their competition for adopting social change and capturing the benefits from innovation. A city participates to this interurban competition through the functions (or economic specialisation) that it successively acquires over time. A function enable a city to supply a type of product or service to other cities, which provide more or less returns in terms of economic growth and

attractivity on population, according to the level of productivity of that function. The criteria for establishing a list of relevant specialisation for the definition of urban functions are related to an evolutionary perspective, under the main hypothesis that it exists a narrow connexion between the relative dynamics of an urban entity in the system of cities and the innovation cycles that the city has adopted (or to which it has better adapted). The question is to identify, for the entire system of cities, which innovation cycles have produced noticeable urban specialisation, affecting in a durable way the relative evolution of the specialised cities, and for each city, which are the specialisation that correspond at best to its actual and potential trajectory. A limited number of urban functions were selected as representative of the major economic cycles that gave rise to differential urban growth and cities specialisation over the past four centuries (Fig. 4). Cities as agents have a total or partial (as constrained by the network of their partner cities) information about the emergence of new functions (which remains exogenous to the model). Cities also have a power of decision to invest in that innovation, according to the wealth they have previously accumulated and to their line of urban strategy that can be more or less risk-oriented. This decision process is represented by a “cognitive” attribute named “urban governance”. The urban governance also may represent in the model the possible intervention of the individual actors, which represent a third level in the modelling of urban systems. This level can be lower than the city level (for instance, an investor choosing a specific location for a firm, or a mayor defining a type of urban policy) or above the system of cities (for instance a political system imposing a centralised administration can lead to the emergence of a prominent capital in an urban system (example of France), whereas a more decentralised government may lead to a more regular urban hierarchy (example of Germany)).

3.2 Simulation of “Second Order” Interactions

What are interurban interactions in the SIMPOP2 model? As our intention is to simulate the development of urban systems that include a large number of towns and cities, it would be unrealistic to think of interactions as “real” flows of exchanged goods, people or information. The interactions which are simulated in the model are not these “first order” interurban exchanges, but more abstract, “second order” interactions, which represent an interpretation of the effect of concrete flows on the relative dynamics of cities. For example, the urban functions are essential attributes of the cities. They do not give an exhaustive representation of the economic profiles of the cities, since they are attributed only to the cities having developed a major specialisation in a particular sector of activity during the corresponding innovation cycle. In a similar way, the exchange of products and resources among cities on the “market place” (cities selling and buying according to their level of supply and demand) does not reflect the totality of the urban economy but only the specialised part of the interurban market, the one that is likely to give rise to urban growth differentials (Fig. 5).

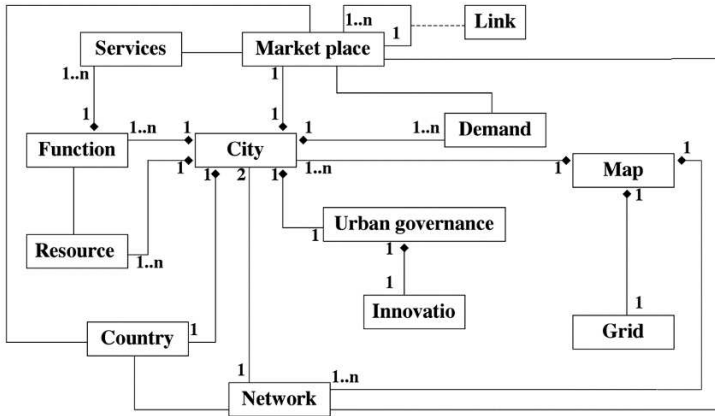


Fig. 5. Main features of SIMPOP2

The rules which define the ability of a city to adopt innovations are partly deterministic, in order to reproduce the powerful trend to hierarchical diffusion of urban innovation (this is the case for most of central functions, a given level cannot be acquired if the other are not yet there), and partly random: when new urban specialisations appear, they can select locations (or become acquired according to some decision of urban governance) which do not necessarily correspond to the largest cities. There are sometimes necessary “seeds” for such location of specialised activities, as mineral resources for manufacturing industries of the 19th century, or knowledge (human capital) in the case of technopoles of the 20th.

3.3 Three Types of Spatial Interaction

Our model is a geographic model, in the sense that spatial interaction are supposed to reflect the power of cities in terms of range of influence of their activities and support for new developments from their access to more or less extended markets. Three types of spatial interactions are distinguished for reflecting the most frequent types of interurban exchanges, linked to different constraints: 1) proximity constrained interactions are representative of many activities for which the distance between supply and demand is an essential constraint, they are the rule for all central place functions, whatever their level and range, and even if that spatial range is increasing over time; under that rule, the probability of exchanges are distributed according to a model of gravity type 2) territorially constrained interactions are limiting a city’s influence within boundaries, regional or national, they correspond to all types of administrative or political activities; the interaction rule is modulated according to the nature of the activity, for instance, a capital can levy taxes in an exhaustive way on all cities belonging to its region or state, whereas in the case of other activities this rule can attribute only a preference for a territorial

market; 3) interactions within specialised networks are free from distance constraints, even if exploring them for developing new markets along this line may have a differential costs according to the distance. Long distance trade, maritime transport, part of tourism activities, or manufacturing industry, are following this type of spatial interaction rule (Fig. 6).

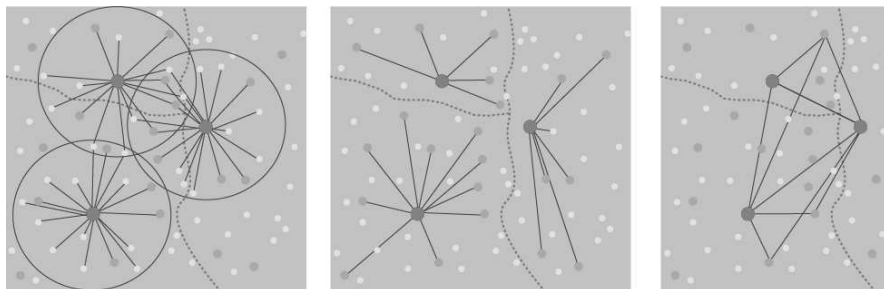


Fig. 6. Three types of spatial urban interactions in SIMPOP2. Left: market; centre: territorial; right: network.

4 Conclusion

Various experiments in simulation are under way for exploring the reliability of our geographical multi-level representation of the dynamics of urban systems. The generic model SIMPOP2 conceived in this way enables the emergence of an urban hierarchy from the interactions between cities, as long as a continuous flow of innovations (composed of new urban functions) is entering the system. We have to check its further capabilities at two levels:

1. reproducing a diversity of plausible growth trajectories at the level of individual towns and cities (which can be interpreted in connexion with their functions, as well as their geographical situation, but also according to the random accidents as simulated by the model).
2. reproducing the emergence of different systems of cities according to the timing and intensity of the urbanisation process, which are produced by rules given exogenously (period of beginning of the urban transition, pace of urban growth during the transition, type of national territorial governance, and so on). This is the type of results to be expected from a meso-macro simulation model using a multi-agents system.

Practical issues could be also investigated by more detailed and adapted versions of the model, as for instance the questions linked with optimal urban development strategies in Europe, either concentrated or polycentric, as well as questions of local urban management within a global context.

In this respect it is necessary to discuss the paradigm of contemporary complex systems theory that focus on the emergence of properties at a macro level as resulting from micro level behaviour. Most of applications, for instance in cognitive economy or in social networks, refer to a “social ontology” which does not include aggregated entities as having an important role in the structuring of social systems. The generation of institutions and their intervention in economic modelling remain as a black hole in economic theory. Even if we think it interesting to complete the theory of evolving urban systems, by relating the observed or simulated features of this growth process to the social processes which are shaping them (which social practices and, if possible, which intentions are behind the observed statistical aggregated properties?) we are not sure that this can in a near future produce a spectacular advance in research. On one hand this view reflects a new fashion in science with a taste for considering bottom-up constructions and sometimes neglecting too much other determinants from aggregate levels (Pumain 2005). On the other, our empirical knowledge about detailed intercity flows (of persons, goods, information, decision-making) and their corresponding motives at the level of individuals are still very limited (Pumain and Racine 1999). So the empirical testing of the hypothesis that will be put in such models has to be very cautious. But in any case, it is important to define this as a promising research program.

References

- Allen P (1991) Spatial models of evolutionary systems. In: Pumain D (ed) *Spatial analysis and population dynamics*. John Libbey-INED, Congresses and Colloquia 6, Paris
- Allen P (1997) *Cities and regions as self-organizing systems: models of complexity*. Gordon and Breach, Amsterdam
- Allen P, Sanglier M (1979) A dynamic model of growth in a central place system. *Geographical Analysis* 2: 256-272
- Andersson C, Hellervik A, Lindgren K, Hagson A, Tornberg J (2003) The urban economy as a scale-free network. ArXiv.cond-mat/030355 v2
- Anderson P, Arrow K, Pines D (eds) (1988) *The Economy as an Evolving Complex system*. Addison Wesley, Redwood City
- Atlan H (1979) *Entre le cristal et la fumée*. Seuil, Paris
- Arthur BW (1994) *Increasing returns and path dependency in the economy*. Michigan University Press
- Arthur WB, Durlauf S, Lane D (eds) (1997) *The Economy as an Evolving Complex System II*. Addison Wesley, Redwood City
- Axtell RL, Florida R (2000) *Emergent cities: a microeconomic explanation of Zip’s Law*. Center on Social and Economic Dynamics, Working Paper, The Brookings Institution
- Batty M, Torrens PM (2001) Modelling complexity. *The limits to prediction*. *Cybergeo* 201: 24
- Berry BJJ (1964) *Cities as systems within systems of cities*. *Papers and Proceedings of the Regional Science Association* 13: 147-163
- Bretagnolle A (1999) *Espace et temps dans l’évolution d’un système de villes*. PhD Thesis, University of Paris I, France

- Bretagnolle A, Mathian H, Pumain D, Rozenblat C (2000) Long-term dynamics of European towns and cities: towards a spatial model of urban growth. *Cybergeo* **131**: 17
- Bretagnolle A, Paulus F, Pumain D (2002) Time and space scales for measuring urban growth. *Cybergeo* **219**
- Bura S, Guerin-Pace F, Mathian H, Pumain D, Sanders L (1996), Multi-agents systems and the dynamics of a settlement system. *Geographical Analysis* **28**, 2: 161-178
- Casti JL (1997) *Would-be worlds*. John Wiley, Chichester
- Casti JL, Swain H (1975) Catastrophe Theory and Urban processes. *Optimization Techniques* **1975**: 388-406
- Clarke GP (ed) (1996) *Microsimulation for Urban and Regional Policy Analysis*. Pion, London
- Drogoul A, Ferber J (1994) Multi-agent simulation as a tool for studying emergent processes in societies. In: N Gilbert, JE Doran (eds) *Simulating societies :the computer simulation of social phenomena*. University of London College Press, London, pp 127-142
- Durlauf SN (2003) *Complexity and empirical economics*. Santa Fe Institute, Working Paper 03-02-013
- Epstein JM, Axtell R (1996) *Growing Artificial Societies: Social science from the Bottom Up*. MIT Press, Cambridge, MA
- Ferber J (1995) *Les systèmes multi-agents: vers une intelligence collective*. InterEditions
- Fujita M, Krugman P, Mori T (1994) On the evolution of hierarchical urban systems. North American meeting of the Regional Science Association, Niagara Falls.
- Fujita M, Krugman P, Venables AJ (1999) *The Spatial Economy*. MIT Press, Cambridge, MA
- Hägerstrand T (1952) The propagation of Innovation waves. *Lund Studies in Geography B* **4**
- Haken H (1977) *Synergetics – An introduction; Non equilibrium Phase Transitions and Self-organisation in Physics, Chemistry and Biology*. Springer, Berlin
- Holm E, Sanders L (2001) Modèles spatiaux de micro-simulation. In: Sanders L (ed) *Modèles en analyse spatiale*. Hermès, Paris, pp 187-217
- Hubermann BA (ed.) (1988) *The ecology of computation*. Elsevier, North Holland
- Langton CG (1988) *Artificial Life*. Addison-Wesley, Reading, MA
- Lesourne J, Orléan A (eds) (1998) *Advances in self-organization and evolutionary economics*. Economica, Paris
- Lewontin RC (2003) Four Complications in Understanding the Evolutionary Process. *Santa Fe Institute Bulletin* **18**: 1: 17-23.
- Livet P (1983) La fascination de l'auto-organisation. In: Dumouchel P, Dupuy J-P (eds) *op.cité*, pp 165-171
- McKinnon D, Cumbers A, Chapman K (2003) Learning, innovation and regional development: a critical appraisal of recent debates. *Progress in Human Geography* **2**: 293-311
- Page SE (1998) On the emergence of cities. Santa Fe Institute, Working Paper 98-08-075
- Page M, Parisel C, Pumain D, Sanders L (2001) Knowledge-based simulation of settlement systems. *Computers, Environment and Urban Systems* **25**: 2: 167-193
- Prigogine I (1996) *La fin des certitudes*. Odile Jacob, Paris
- Pumain D (1998) Urban Research and Complexity. In: Bertuglia CS, Bianchi G, Mela A (eds) *The City and its Sciences*. Physica Verlag, Heidelberg , pp 323-362
- Pumain D (2000) Settlement systems in the evolution. *Geografiska Annaler* **82B**: 2: 73-87
- Pumain D (2003) Une approche de la complexité en géographie. *Géocarrefour*

- Pumain D (ed) (2005) *Hierarchy in natural and social sciences*. Kluwer, Dordrecht
- Pumain D, Racine JB (1999) L'homo geographicus: de l'emprunt à l'empreinte. *Revue européenne des sciences sociales* **113**: 77-86
- Robic MC (1982) Cent ans avant Christaller: une théorie des lieux centraux. *L'Espace Géographique* **11**: 5-12
- Robson B (1973) *Urban growth, an approach*. Methuen, London
- Sanders L (1992) *Système de villes et synergie*. Anthropos, Paris
- Sanders L, Pumain D, Mathian H, Guérin-Pace F, Bura S (1997) SIMPOP, a multi-agents system for the study of urbanism. *Environment and Planning B* **24**: 287-305
- Schweitzer F (1998) Oekonomische Geographie: Räumliche Selbstorganisation in der Standortverteilung In: F Schweitzer, G Silverberg (eds) *Evolution und Selbstorganisation in der Oekonomie*. Duncker & Humbolt, Berlin, pp 97-125
- Taylor PJ (1997) Hierarchical tendencies amongst world cities. *Cities* **14**: 6: 323-332
- Tesfatsion L (2002) Agent-Based Computational Economics : Growing Economies from the Bottom Up. *Artificial Life* **8**: 1: 55-82
- Weidlich W (2000) *Sociodynamics*. Taylor and Francis, London
- Weidlich W, Haag G (1988) *Interregional migration – Dynamic Theory and Comparative Analysis*. Springer, Berlin
- Wilson A (1980) *Catastrophe theory and bifurcation*. Croom Helm, London
- Wolfram S (1994) *Cellular automata and complexity*. Addison-Wesley, Reading, MA
- Zelinski W (1971) The Hypothesis of the Mobility Transition. *Geographical review* **61**: 2: 219-249

The Great Return of Large Scale Urban Models: Revival or "Renaissance"?

Giovanni A. Rabino

Department of Architecture and Planning (DiAP), Polytechnic of Milan, Italy

Abstract

There are many evidences for a great return of large-scale urban models. Nevertheless this revival could be a true renaissance, only if the modellers will win seven challenges, in dealing with changes in urban phenomena, in urban theories, in planning theories and practices, in modelling activity, in types of models and in technical aspect of modelling. The supreme challenge, however, in building good urban models will be to join scientific and classical cultures.

1 Introduction

Operational Large-Scale Urban Models (OLSUMs) are mathematical simulation models describing urban systems in great details (spatial, with a fine zoning of the territory, and functional, with a disaggregated account of urban activities and infrastructures) for applications in planning and policy practices.

This kind of modelling begins, in USA, in late '50s of the last century; and flourishes at best in '60s and '70s, especially in UK in a large number of applications for sub-regional territorial planning (Batty 1976). These models lose much of interest in '80s and '90s, when only a restricted set of scholars carried on the modelling tradition with relatively few, but nevertheless very important, applications (Wegener 1994). On the contrary, since the beginning of the new millennium a slow but growing interest can be remarked, as pointed out also by other authors (Torrens and O'Sullivan 2001; Brail and Klosterman 2001). This entire story is briefly described in the following section 2.

The question is: why have OLSUMs such a strong return of interest? In section 3 we will trace the causes exactly in failing of those factors that were the roots of the models' decline in '80s. Moreover, beyond a simple though vast and prominent revival, the author points out a good chance of a true "Renaissance" of urban modelling, if some opportunities will not be missing.

To catch this potential means to be able to not forget in building models the changes occurred since the '60s in society (then in cities), in science and technology (then in modelling activity itself). Thus six challenges to modellers, namely to their care of these changes and skill in dealing with, are presented in section 4.

As seventh -supreme- challenge, the moral of this paper is drawn in the closing section 5.

2 Evidences of OLSUMs' Return

To defend the thesis of a meaningful return of large-scale urban modelling (section 2.2), it is worthwhile telling, at least shortly, the history of that kind of models. It will be done in the next section. By the way, this lets you, my dear reader, learn (or, perhaps for someone, remember) something about that "golden age" of urban modelling.

2.1 A Very Brief History of Operational Urban Modelling

Birth and growth of OLSUMs has been triggered, in the '60s, by the co-occurrence of three factors:

- New problems: the need of understanding and managing the USA metropolis, now based on (and crowded by) heavy commuters' flows, when traditional master plan becomes ineffective;
- New theories: development at that time, or a bit earlier, of a consistent *corpus* of theories¹ good for facing the new problems;
- New computation tools: computers start to become usable for non-military applications, and their computing power² allows the practical application of the new theories, *via* numerical solution.

Beyond the peculiarities of each model, all they share a common mathematical structure, of input-output type:

$$X = (I - A)^{-1}Y \quad (1)$$

where the elements:

x_{im} of the matrix X is the level of urban activity (or stock) m in zone i ;

y_{jn} of the matrix Y is the level of some exogenous demand (from the outside of the town) of urban activity (or stock) n in zone j ;

a_{imjn} of the tensor A is a coefficient measuring the spatial and functional dependence of activity i in zone m from activity n in zone j (land-use constraints included). This coefficient usually is a function of X and Y .

Non-linearities in A and inversion of $(I - A)$ need an iterative solution of Eq. 1.

¹ Such as the theories and methods of regional science (Isard 1960) or the dynamic systems theory (Forrester 1968)

² An incredible power for that time, even if it seems ridiculous nowadays. E.g.: a Lowry model (see later) of the '60s can be solved today by a macro of an usual spreadsheet (<http://people.hofstra.edu/geotrans/eng/ch6en/meth6en/ch6m2en.html>)

A list of those OLSUMs includes³:

- Greensborough (linear econometrics) model, by Chapin and Weiss;
- EMPIRIC model of the Boston Region, by Hill;
- Baltimore and Connecticut models, by Lakshmanan;
- Delaware Valley (nonlinear activities allocation) model, by Seidman;
- Bay Area Projective Land Use Model (PLUM) (gravity model), by Goldner;
- Upper New York State model, by Lathrop and Hamburg;
- Penn-Jersey model (mathematical programming for optimal allocation) by Herbert and Stevens and subsequently by Harris;
- National Bureau of Economic Research (NBER) econometric model for urban simulation, by Ingham, Kain and Ginn;
- South East Wisconsin Land Use Plan Design models, by Schlager;
- Bay Area Simulation Study (BASS) (mixed techniques model), by Wendt et al.;
- San Francisco Housing Market Model, by Robinson, Wolfe, and Darringer.

But a special mention is due to the Pittsburgh model (Lowry 1964), because of its enormous influence over all the history of large-scale urban models; firstly, the so-called “*the Lowry heritage*” (Goldner 1971), a wealth of Lowry type (but largely improved) models applied in British sub-regional plans (Bedford, Lancashire, Nottingham-Derby, Cheshire, Reading, ...) ^{4 5}.

In the ‘60s, finally, we can’t forget two other kinds of models, based on different mathematical approaches:

- Forrester’s Urban Dynamics model (and followers), a set of coupled first order difference equations, each one related to changes in level of an urban activity (Forrester 1969);
- Community Land Use Game (CLUG), by Feldt (1972), and other similar toys (till to SimCity, a bit later), game simulators intended for training of planners and remarkably based on same rules and relations of Lowry model.

For the reasons we recall below (section 3.1), OLSUMs production went greatly down in ‘80s and ‘90s. Wegener (1994), in an excellent “state of the art” report, makes a list of only 20 (more or less) active urban modelling centres all over the world: Philadelphia (Putman); Cambridge (Echenique); Melbourne (Brotchie and Roy); Stockholm (Lundqvist and Mattson); Buffalo and Chicago (Anas, Boyce and Batty); Caracas (de la Barra) and Dortmund (Wegener), just to mention someone. He lists also few (no more than twelve) models, developed in

³ There are some excellent reviews. As a starting point for the reader we suggest Batty (1976).

⁴ Let me remember here an Italian application too (the model for the subregional plan of Biella, by Bertuglia and Rabino 1975)

⁵ A great contribution to define theoretical bases and to organize in a systematic way all this kind of modelling activity came from Wilson (1974) and co-workers in Leeds.

the centres. Of those models⁶ the well-known ones were, in my personal judgement:

- MEPLAN, the Echenique and partners' model, surely the most successful OLSUM in every time, applied to two dozen of urban regions, or more (Cambridge, London, Santiago, Sao Paulo, Bilbao, Naples, Helsinki, Tokyo, ...);
- TRANUS, de la Barra's model, applied especially in Latin American cities;
- IRPUD, the model firstly developed by Wegener for Dortmund, and than used for other regions, such as to study the European spatial organization;
- ITLUP, a second success story of urban modelling, the model of Putman applied in planning of dozens of U.S.A. Metropolitan areas;
- TOPAZ, "optimal allocation of activities" model built at CSIRO (Melbourne) by Brotchie et al.

Like in the '60s models, but now with much emphasis, a common feature of these models is comprehensiveness: within a framework of input-output type⁷, they aim to take in account all the aspects (and relations) of an urban environment: population, housing, land-use, workplace, employment, network, travel and goods transport.

Two important characteristics peculiar to '80s and '90s urban models are:

- the focus on transport analysis and planning. So these models are sometime called "integrated land-use transport models";
- a special attention paid to sound theoretical (economic) foundations, such as random utility and locational surplus in choice processes, bid-rent in pricing mechanisms, equilibrium principles in land-use and/or transport assignments.

After all, to give a full account of the story, we must quote the "quasi-operational"⁸ urban models:

- in '80s a few prototypes of really dynamic (i.e., evolutionary) urban models⁹;
- in '90s some attempts to build models using new A.I. methods¹⁰.

⁶ For a review see also: Oryani and Harris (1996)

⁷ Sometime in a single "unified" model; usually "composite" models, a set or more or less loosely coupled submodels (e.g. ITLUP, made by DRAM land-use model plus EMPAL transport model).

⁸ "Quasi" because of lack of suitable time-series of data for dynamic models; and because early rough prototypes of the A.I. models.

⁹ Interacting logistics approach (Allen and Sanglier 1979; Pumain et al. 1989); quasi-equilibrium approach (Lombardo and Rabino 1984); multi-state approach (Haag 1989)

¹⁰ Such as neural networks (Fischer and Gopal 1994), cellular automata (White et al. 1997) and evolutionary methods (Openshaw 1988)

2.2 OLSUMs of the New Millennium

Surfing the Internet, that is nowadays the same as doing a traditional bibliographic survey, we discover probably more (websites of) urban models than all the models of '60s golden age.

A surely incomplete list includes¹¹:

- the evergreen MEPLAN¹² and TRANUS¹³, already mentioned in the '80s models;
- the family of CUF (California Urban Future) models (i.e. CUF, CUF II and CURBA)¹⁴ developed by Landis at University of California (Berkeley) that were first models using a GIS software for managing data, computing and presenting results;
- METROPILUS¹⁵, restyling (with some improvements) of Putman's ITLUP, now a model with a easy-to-use graphical user interface and GIS integrated;
- URBANSIM¹⁶, developed by Waddell at Washington University (Seattle). Theoretical bases and conceptual structure of this model are the usual ones of the '80s models. However an operational feature needs a note: a flexible (object-oriented) software architecture, to make components reusable in future developments in the modelling activity;
- SLUETH¹⁷, a land-use model based on cellular automata approach (land-use transitions are functions of slope of terrain, proximity to roads, diffusive growth, spread of new centres and physical and/or planning constraints). In Project Gigalopolis, Clarke, the model developer, is in search of a general relation between the values of these variables and spatial patterns, by applications to a worldwide set of urban case-studies;
- WHAT IF, COMMUNITY-VIZ, INDEX, METROQUEST¹⁸ are all Planning Support Systems, including in their structure a (more or less complete and advanced) urban modelling module. Being PSS (sometime to be used for public participation in planning), they have good facilities for handling the urban model: user-friendly interfaces, GIS and 3D visualization, easy interoperability with other software (e.g. commercial transport models), evaluation indices.

¹¹ The list is an extension of the survey reported in Brail and Klosterman (2001)

¹² Product of the M. Echenique And Partners Limited (Me&P), now acquired by the Swedish WSP Group (<http://www.wspgroup.com/uk>)

¹³ de la Barra model can be downloaded from website: <http://www.modelistica.com/>

¹⁴ CUF focuses on housing and residential locations; CUF II takes in account all socio-economic aspects; CURBA is the first model that adds environmental considerations (habitat conservation) to urban modelling

¹⁵ http://gis.kent.edu/gis/empact/lit_urb_md01.htm

¹⁶ <http://www.urbansim.org>

¹⁷ <http://www.ncgia.ucsb.edu/projects/gig>

¹⁸ <http://www.what-if-pss.com>, <http://www.communityviz.com>, <http://www.crit.com>, <http://www.envisiontools.com>

3 The Core Question: Revival or “Renaissance”

To explain the fall and resurrection of urban modelling, it is useful firstly to look into the causes of OLSUMs decline in the ‘80s (next section 3.1). Then, on these bases, we can argue the present revival and speculate about a possible “Renaissance” of these models (i.e., not just a return but a really bright “new era” of their history). Section 3.2 presents the arguments.

3.1 Seven Sins of ‘60s Large-Scale Models

Lee (1973), in his very famous paper “Requiem for large-scale models”, sentenced them to disappear because of seven deadly sins; and later all the scholars of modelling agreed with his analysis¹⁹.

The sins were:

1. Hypercomprehensiveness. OLSUMs can be seen as the operational counterpart of the so-called rational Comprehensive Planning (McLoughlin, 1969; Chadwick, 1970). Weaknesses of this mode of planning (too strong top-down control, too broad scope and too many purposes) are weaknesses of the models;
2. Grossness. Despite of the usual very disaggregated analysis, in those models relations among variables were stated at macro-scale level²⁰; so often the model is unable to meet the detailed description needed for planning purposes;
3. Hungriness. Today like in those times, urban models call for a lot of very detailed information as inputs; yet the data are sometime absolutely unavailable²¹, often very hard to find or costly to collect;
4. Wrongheadedness. The concern is with the merely descriptive²² (of a given situation) nature of many of those models. Accordingly they are wrong in doing forecast when something changes;
5. Complicatedness. The “complex” view of things, peculiar to the (new at that time) “systems theory” approach, was not immediately accepted by ordinary people (such as practitioners or decision makers) and even by some Academic *milieu*. Large-scale models seemed to be complicated and incomprehensible, instead of rightly complex;
6. Mechanicalness. Decoding the Lee’s statement²³, the criticism is for too much rigid and deterministic input-output relations (in the models), forgetting the fuzziness of human (and therefore urban) behaviours;

¹⁹ See for instance Harris (1994)

²⁰ Frequently statistical relations, good for an enough large set of elements

²¹ In this case usually “proxy” are used for those variables

²² In other word, they don’t explain the facts upon theoretical bases

²³ The real concern of Lee is with the errors, such as rounding errors, in automatic computations; and the impossibility to estimate them. In my opinion, this is only a side of the more general problem

7. Expensiveness. In the '60s, to build and apply an OLSUM needed three, four or more years²⁴ (without being sure of a positive outcome) of work of large research teams. Thus costs (financial and time costs, engagement of people and use of material resources) overcame a lot the model benefits, with only few public administrations able to afford this innovative support to planning.

3.2 A Sure Revival and a Possible "New Era" of Urban Modelling

Nowadays, looking at the past, it is possible to assert that large-scale modelling in the '60s has been an ambitious experiment, probably too much ahead for its time.

Changes in society, science and technology, especially in computer science and tech, are going to forgive (or already forgave) the sins:

- hypercomprehensiveness. Certainly model features and planning style are tightly connected (see section 4.3), but on the bases of this knowledge we are able now to build a model suitable for a given planning approach;
- grossness. We are skilful nowadays in using the macro, meso or micro level of description most appropriate to deal with a given planning problem;
- hungriness of data may be still a problem. Nevertheless increasing availability of automated and remote sensors, and extraordinary diffusion of large computerized data-bases²⁵ are going to set a reverse situation of extreme data supply²⁶ (perhaps with some new problems);
- wrongheadedness. This was a Lee's overstatement, effective only for few "statistical" models. Huge developments in urban theories, from '50s till now, make the criticism totally pointless;
- complicatedness. Complexity is not a problem any more; perhaps it is too much a fashion nowadays;
- mechanicalness. Today new urban theories (see next section 4.2 for details) let us build models, envisaging irreversible, counter-intuitive and many other qualitative aspects of the vagueness of urban phenomena²⁷;
- expensiveness. Since the '60s the costs are falling, slowly but steadily, making modelling affordable by an increasing number of municipalities, even because of a greater awareness of benefits of these planning supports by public administrators.

Being sure for these reasons of an important revival of OLSUMs, always on the bases of the above-mentioned changes, plus some forecasts easy to do, I am won-

²⁴ A few examples (approx. dates): Lowry, 1961-1965; EMPIRIC, 1964-1968; PLUM, 1967-1970; Greensborough model, 1960-1966

²⁵ I mean, firstly, GIS and on-line archives

²⁶ On this point see, for instance, Lutzak (2005)

²⁷ Evolution of informatics, more powerful computers and a greater trust in automation overcome the Lee's doubts about errors, too.

dering if urban modelling could flourish, in a “Renaissance” new era, by taking into account more explicitly in model building those changes and previsions.

4 Six Challenges (Plus One) for a Renaissance of Models

My hope, for a bright near future of OLSUMs, leads to challenge urban modellers’ talent in dealing with six changes (see sections immediately below), in their modelling activity. A seventh challenge, an overall challenge more tricky to tackle, is then presented in the next final paragraph.

4.1 Changes in Urban Systems

In the late ‘50s, when urban modelling began, only one third of the world population (of about 2.5 billion people) was living in cities; and for all the people, or at least for the modellers, the metropolis *par excellence* was the Fordist town (industrial towns like Detroit, Manchester, Turin, ...). All the models were conceived for that kind of city only. Now, at the end of the 20th century, when one half of about 6.0 billion people inhabits the world towns, we (and modellers) are much more aware of the multitude of urban settlements types. Following the Soja (2000) example, the variety of cases can be classified according to a series of discourses about:

- the *flexible* town. New information and communication technologies (ICT) are re-shaping (goods and services) production²⁸ and the labour organization²⁹. All these changes are upsetting the infrastructural (places and networks) and functional (activities and flows) structure of the cities. But they are taking place in peculiar ways and with unequal speed in different cities. Models must account each specific situation;
- the *glocal*³⁰ town. On one hand a town is a metropolis of the world³¹; on the other hand, it is a unique entity, for its history, culture, traditional productions, and so on³². Models must accordingly try to deal with this (*sic*) complementary clash;

²⁸ To give an idea of these changes, it is enough a mention to new informatic services, just-in-time products “for you” tailored, “virtual” factories,

²⁹ It is enough an hint to new professions, tele-working, flexible labour market,

³⁰ A neologism made by global plus local terms

³¹ It is no more just a central place of a region or a country. For financial, social or economic reasons it is a node of one of the world wide networks, competing or making alliance with other towns ...

³² These dia-logic (but each other explanatory) views of a city are a very good example of urban complexity (see section 4.2).

- the *polymorphic* town. Classic Thünian mechanism, leading to a “compact” city³³, no more holds alone. We³⁴ are faced now with a range of locational processes on demand side and a lot of supplying works. Both are highlighted by the variety of urban patterns: megalopolis, edge cities, urban sprawl, urban areas, exurbanization, ... ;
- the *plural* town. The great deal of new and old “social” aspects of the cities can’t be neglected in modelling. Few examples are: spatial segregation (slums, gentrification ...), multi-ethnic population (cultural integration, economic divide ...), lifestyles (new kinds of families and other social aggregation, elderly people and children, time organization ...);
- the *risky* town. There is an increasing demand by people of security in towns. And this risk aversion covers all kinds of fear: natural risks (such as river overflowing), industrial risks and like those (plant explosions, fires in building, electric black out, shortage of goods, ...), health risks (contagious diseases, car accidents ...) and crime risks (both the ordinary criminality and exceptional events, like terrorist attacks). For policies and planning purposes³⁵ models must include those aspects;
- the *virtual* town. In the next section we will outline the cities as artefacts of the human minds, before manual labour. The world of information (all kinds of media, but especially TV and internet) plays a key role in this great game. Thus there is now a second city, a virtual “city of information” parallel (but strongly interconnected) to the real one; and this “e-city” (teleworking, e-commerce, e-government, ...) is probably as much important as the other. This city will necessarily be a part of a comprehensive urban model;
- the *ecological* town. Environmental problems are a major concern in our society. And these problems are mainly “urban” environmental problems. So urban models must deal with aspects like the sustainability of the urban development, the function of natural ecosystem in the cities, and so on.

The challenge is to build for a given town a specific model dealing with the peculiarities of that town with reference to all the above-mentioned discourses.

4.2 Changes in the Urban Theories

Compared to the ‘60s and ‘70s, there has been a major change in science, the emergence of the so-called “science of complexity” (in natural and artificial sciences³⁶ as well as in classical studies³⁷).

³³ Urban density declining with distance from centre

³⁴ I.e. urban theories and models

³⁵ For instance to study “domino” effect in catastrophes

³⁶ See for instance the physicist Haken (1997) or the artificial science expert Wolfram (2002)

³⁷ See for instance the sociologist Morin (1990) or the philosopher (and chemist Nobel laureate) Pigroine (1980)

This new scientific approach is re-defining the disciplines, making the “classic” one just an approximation or a special case; and this revolution is now active especially in tricky sciences in between natural and human studies, such as urban studies (i.e. analysis and planning sciences).

In my view³⁸, there are three aspects of urban complexity that theories and models must tackle:

- Synergies. These are the most common phenomena regarded when people thinks about complexity. I mean: non-linearity, irreversibility, discontinuity, chaotic and other strange behaviours, power law, random networks, fractals All these aspects are promising fields or research³⁹ and all together contribute to develop a new view (i.e. theory and modelling approach) of urban systems: the evolutionary town⁴⁰.
A note is worthwhile: the experience acquired in applying quasi-operational (or merely experimental) "dynamic urban models" in the '80s and '90s (see end of section 2.1) may be very fruitful in development of the new evolutionary OLSUMs;
- Qualitative aspects of towns⁴¹. Things are qualitative (or just lexical) when we are unable to define them in a more rigorous way because of underlying complexity. Examples are ethic rules, aesthetic values, collective and personal believes, experiences and hopes. Clearly, all this amount of information concerning the city (surely larger then all quantitative data) must be embedded in models. And now we are going to have tools to do that⁴².
- A cognitive view of the city⁴³. A unique feature of urban systems is that they are made by conscious people able to think over the city itself⁴⁴. If the OLSUMs of the first generation had the great merit of rightly unifying several disciplines (land-use planning, economy, transport...), this new cognitive approach must be credited for widening the analytical focus to consider that cities derive (maybe, mostly) from “human culture”. Such an approach, careful to the aspects of knowledge and intentionality of the human actions, can’t be neglected any more in urban modelling⁴⁵.

³⁸ My interpretation of “science of complexity” as the re-joining of analysis and synthesis (or, to be more precise, design of the system) in systems studies, after a long time of unjustified antithesis is explained in Rabino (2005)

³⁹ There are already many ongoing studies

⁴⁰ Let me say that I was probably the first to highlight this view (Rabino 1993)

⁴¹ I pay a special attention to this point, because of my education in humanities

⁴² For instance “cross impact analysis” is a method for modelling systems described in qualitative terms (Blecic et al. 2005)

⁴³ A special attention to this question is paid by Occelli (2002) and Portugali (2000).

⁴⁴ Note that this fact implies philosophic problems (related to the subject – object relation in scientific studies) and possible paradoxes (such as the self-fulfilling prophecies).

⁴⁵ Being our focus on human agents, “multi-agent systems” (Ferber 1999) seems to be a very appropriate tool for modelling.

4.3 Changes in the Planning Theories and Practices

The models must be especially designed for the specific scopes (in our case, usually for planning purposes) for which they are built⁴⁶. From the hierarchical top-down, rational-comprehensive plan⁴⁷ of the '60s, now the theoretical positions about planning styles have been enriched and diversified⁴⁸ (according three directions of reading):

- in the real nature of the plan. Looking at planning as a problem setting and solving process⁴⁹, the core of this activity can be more or less centred on the object, on decision, on action, on communication⁵⁰...;
- in the "role" of the planner. Obviously, the features of a plan depend on the meaning that the planner (a single or a group of practitioners) attaches to his work. There are planners: a) with a more romantic view or a more rational outlook on their activity; b) more attentive to theories or more skilled by experience.

⁴⁶ Two simple examples are:

- a "cross impact analysis" very useful in a collaborative building of scenarios, but totally useless for quantitative evaluations;
- a "land-use cellular automata" good for forecasting urban sprawl, but of little utility in settlement of policies concerning urban spaces.

However the actual cases are more complex of these two examples; and they need a huge work for tuning the model to the circumstances.

⁴⁷ That is a planning practice based – in short – on two assumptions:

- the goals of the plan must be defined moving from the larger to the smaller ones (ie. from national level to local scale); and the goals are fixed by public administrators, on the bases of their view about people's needs;
- the plan is the outcome of a fully logical analysis of the situation, able to solve the problems; and this ability comes out from a system view of the whole case-study, besides the classical rationality.

⁴⁸ This variety derives from a history of planning (Carta 2003) briefly summarized in five phases:

- until '50s, plan as a more or less technical design;
- 60s, plan as a systemic "visioning" of a preferable future;
- 70s, plan as a political action (decisions about the future);
- 80s, focus on putting the plan into effect;
- from '90s till now, focus on means (such as public participation, reasoning, ...) of promoting the accomplishment of the plan.

⁴⁹ From perception of the problems till to management of solutions through solutions finding, comparative evaluations, decision making, implementation (involvement of human and economic resources, laws and rules making, etc.), conduct of the action ...

⁵⁰ That are, respectively, the focuses of the so called systemic, political, transactive and participative plans.

Combining these two aspects, Udy (1991) classifies four types of planners: the administrator, the promoter, the manager, the reformer⁵¹;

- in the purpose of the plan. Plans usually have a large set of targets, but generally as well these goals can be derived from an overall aim peculiar to the case-study. A list of (well-known to everyone) purposes includes: the engagement for a sustainable development, an enhanced use of new technologies, the fulfilment of social needs, the promotion of a local historical identity in a global economy, and so on.

Beyond this theoretical variety, we can't neglect the lessons⁵² from multiple real practices. Nowadays we are experienced in benchmarking of "best practices", in public participation (management, effectiveness, costs, etc.), in attempts of e-governance (implementation, benefits, social divide, etc.) ... so let me repeat once more: modelling must be embedded in planning practices in agreement.

4.4 Changes in Modelling Activity

It is a long time (see for instance Occelli and Rabino 2006) that urban modellers are pondering on ontology and working of their own activity.

They are acquainted now with the complexity of the modelling process made from two loops of reasoning centred around the *observable*, the phenomena in the geographical world, which are subject to investigation:

1. the internal loop, or "formal building" of the model, referring to the conventional steps underlying a process of abstraction (i.e. modelling). This has its roots in the assumptions held in the mainstream of socio-economic sciences. It proceeds from observation to the formulation of concepts and more formal models of it (the so called *encoding process*), and terminates referring back these latter to the observed reality (through the so called *decoding process*);
2. the external loop, or "substantial building" of the model, that refers to the historical and socio-cultural domain in which the process of abstraction/modelling occurs. It reminds us that modelling activity has to confront itself with the urban issues it aims to address, as well as with the socio-cultural context in which it is developed⁵³.

⁵¹ Within this frame, taking into account also the power of his action (more or less proactive behaviour of the planner and more or less resilience of the system), Udy (1991) comes to define up to sixteen kinds of planners: the bureaucrat, the designer, the philosopher, the activist, the educator, the entrepreneur

⁵² The "learning by doing" an almost infinite set of plan, everywhere in the world since the '50s.

⁵³ By the way, challenges 1, 3 and 4 presented in this paper are inspired by the awareness of this "substantial" modelling activity. The other ones (2, 5 and 6) are connected to "formal" modelling. The 7th challenge is a bridge between these two activities.

Beyond this analysis of modelling process, as to its real nature, a broader view of the concept of modelling is called for, which has been called the *structural-cognitive* shift (Ocelli and Rabino 2000), to indicate a shift:

1. from a view where modelling is an activity through which an understanding of the organizational structure of an urban system is obtained (the structural perspective). According to this view a model is a (simplified) representation of urban phenomena and the ways they are produced;
2. to a view where modelling is an activity for testing, exploring, creating and communicating knowledge about certain urban phenomena (the cognitive perspective). Models therefore are means for representing the working of our knowledge hypotheses (and of their outcome).

So the structural-cognitive shift affects the scope, role and functions of the modelling activity, influencing the two above fundamental loops too. To suggest a label for this novel definition of modelling, a model can be understood as an "agent"; in particular, it may be defined as an ALC agent capable of:

- a. performing a certain course of *Action*, thus enabling the realization of a certain project of investigation of spatial phenomena. This most directly involves the relationships between the syntactic (the methodological aspects of modelling) and representational (the sense associated with the representations of the urban phenomena) components of modelling;
- b. enabling users with a certain *Learning* ability, thus generating stimuli in critical revising both the external and internal loops of the modelling activity likely to trigger new investigation quests;
- c. *Communicate* with other kinds of agents (other models or people), thereby affecting them and/or modifying itself in the process.

The new OLSUMs must be conceived in a ALC agent perspective.

4.5 Changes in Models

Among the challenges for a "renaissance" of the OLSUMs this is perhaps the simplest one (to win). A great amount of new, and old, modelling methods is already in the toolbox of the modeller (Cellular Automata, Neural Networks, Multi-Agent Systems, Genetic Algorithms, Bayesian Network...⁵⁴) and it is sure that

⁵⁴ Don't worry! A full list of techniques exceeds the scope of this paper. It needs a book or at least another long paper, but a reasoned framework of all the mathematical and computerized methods involved in territorial planning since the '50s can be found in Rabino (2007).

others tools will come soon from the A.I.⁵⁵ and from the creativity of scholars (Batty 2005).

The problem here is not the models availability, but the choice of the most suitable model (or mix of models) for a given case study. We must move from a still widely accidental use (dictated by personal pleasure of the modeler and by the fashion of that time) to a "reasoned" application, based on comparative evaluation (empirical and theoretical), on efficient integrations among different tools, on a careful targeting of the model to its aims.

4.6 Change in Technical Aspects of Modelling

To build and apply an urban model is not just writing a set of equations (or an algorithm) and implementing it in a software (with the hope that it runs).

It is a much more complex process involving:

- the use of several other models or computerized tools⁵⁶;
- protocol of action for handling (usually in large working groups of researchers and technician) all the above mentioned machinery and for communicating the results to other planners, to decision makers, to population.

Discussed (only a little and) only in scientific publications, all these technical⁵⁷ aspects of modelling are often managed by the modellers in unprofessional way (i.e. as "bricoleur").

On the contrary, for user-friendly professional OLSUMs applications, with limited costs⁵⁸ and efficiency in communication of the results⁵⁹, it is necessary "to engineer" this activity:

- firstly, with an extensive integration of all the computerized tools involved in modelling (Batty and Longley 2003): on the input side, with automatic and telematic information retrieval systems⁶⁰; on the output side, with other models applied in the planning activity⁶¹;

⁵⁵ See for instance the new frontier of computers "scientist and planner", no more (more or less advanced) tools but true co-protagonists in research and profession, wiping out totally the man-machine divide (Microsoft 2005).

⁵⁶ Such as:

- techniques for data finding; management of data-bases and GIS; methods of pre-processing of input data ...;
- techniques for parameters calibration and fitting of the model to observed data; data transfer from a model to another one ...;
- post-processing of the model outputs; publishing of results

⁵⁷ In the (erroneous N.d.A) sense of merely methodological problems in applying models.

⁵⁸ At least to avoid the reinventing of the wheel every time.

⁵⁹ That are what really interest the decision makers.

⁶⁰ I.e. the remote sensing technologies.

⁶¹ Such as P.P.S. (planning support systems), multi-criterial evaluation tools, or software for 3D visualization of the model results.

- secondly, sharing the “expert knowledge in modelling” and fixing a set of rules (good practices – possibly – codified in a manual) in order to operate at best the above integration.

5 Conclusion: the 7th Challenge

Every occasion of development and application of an urban model is a fascinating adventure,⁶² for the scientist or the practitioner; or, better, for a person (aiming to be) professional and researcher at the same time.

For his work, on one hand he must have a great classical culture, as:

- a sociologist, to realize people’s needs, expectations ...;
- a historian, to recall events, traditions and customs of the town;
- a political expert, to know how decision-making processes work there and to manage his planning activity accordingly;

and so on.

On the other hand he must be gifted with technological skill and scientific culture, in order to be able:

- in stating the working mechanism of the town (in its economic, infrastructural, environmental aspects);
- in dealing with mathematical tools for modelling (looking for the most appropriate one in his case-study);
- in using methods and technologies for the model application (from information retrieval to results exposition);

and so on.

Moreover, he must try to join both competences (often against the prejudices of a culture *vs.* the other one) for connecting at the best the two loops of the modelling process (remember section 4.4):

- the decoding loop, that is well understanding urban features and problems to catch them in the model;
- the encoding loop, that is properly “writing down the formulas” to (help to) find a solution for those problems.

In conclusion, to be a “third culture” (wo)man, in the above sense i.e. according to the dream of Snow (1993), is not just the seventh but the supreme, exciting, challenge for an urban modeller.

⁶² This is a firm conviction of the author, based on his lifelong experience.

References

- Allen P, Sanglier M (1979) A Dynamic Model of Growth in a Central Place System. *Geographical Analysis* **11**: 256-272
- Batty M (1976) *Urban Modelling: Algorithms, Calibrations, Predictions*. Cambridge University Press, Cambridge
- Batty M (2005) *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. MIT Press, Cambridge, MA
- Batty M, Longley PA (2003) *Advanced Spatial Analysis: The CASA Book of GIS*, ESRI Press, Redlands, CA
- Bertuglia CS, Rabino GA (1975) *Modello per l'Organizzazione di un Comprensorio*. Guida editori, Naples
- Blecic I, Cecchini A, Trunfio GA (2005) A Decision Support Tool Coupling a Causal Model and a Multi-objective Genetic Algorithm. *Lecture Notes in Artificial Intelligence* **3533**: 628-637
- Brail RK, Klosterman RE (2001) *Planning Support Systems*. ESRI Press, Redlands, CA
- Carta M (2003) *Teorie della pianificazione: questioni, paradigmi e progetto*. Palumbo Editore, Palermo
- Chadwick GF (1970) *A Systems View of Planning*. Pergamon Press, Oxford.
- Feldt AG (1972) *CLUG: Community Land Use Game*. The Free Press, New York
- Ferber J (1999) Multi-agent systems. Addison-Wesley, Harlow
- Fischer MM, Gopal S (1994) Artificial neural networks: a new approach to modelling interregional telecommunication flows. *Journal of Regional Science* **34**: 4 :503-527
- Forrester JW (1968) *Principles of Systems*. Wright-Allen Press, Cambridge, MA
- Forrester JW (1969) *Urban Dynamics*. MIT Press, Cambridge, MA
- Goldner W (1971) The Lowry Model Heritage. *Journal of the American Institute of Planners* **37**: 2: 100-109
- Haag G (1989) *Dynamic Decision Theory: Applications to Urban and Regional Topics*. Kluwer, London
- Haken H (1977) *Synergetics*. Springer, Berlin
- Harris B (1994) The Real Issues Concerning Lee's "Requiem". *Journal of the American Planning Association* **60**: 1: 31-34
- Isard W (1960) *Methods of Regional Analysis: an Introduction to Regional Science*. MIT Press, Cambridge, MA
- Lee DB (1973) Requiem for Large-scale Models. *Journal of the American Institute of Planners* **39**: 2: 163-178
- Lombardo ST, Rabino GA (1984) Nonlinear Dynamic Models for Spatial Interaction: the Results of Some Empirical Experiments. *Papers of the Regional Science Association* **55**: 83-101
- Lowry IS (1964) *A Model of Metropolis*. Rand Corporation, Santa Monica, CA
- Lutzak E (2005) Extreme Data: Rethinking the "I" in IT. *Leading Edge Forum, Computer Science Corporation* (<http://csc.com>), pp 1-60
- McLoughlin JB (1969) *Urban and Regional Planning: A Systems Approach*. Praeger, New York
- Microsoft (2005) *Towards 2020 Science* (<http://research.microsoft.com/towards2020science>)
- Morin E (1990) *Introduction à la Pensée Complexe*. E.S.F. Editeur, Paris

- Occelli S. (2002) Facing Urban Complexity: Towards Cognitive Modelling. Part 1: Modelling as a Cognitive Mediator. *Cybergeo* (<http://www.cybergeo.fr>)
- Occelli S, Rabino GA (2000) Rationality and Creativity in Urban Modelling. *Urbanistica* **113**: 25-27.
- Occelli S, Rabino GA (2006) Cognitive modeling of urban complexity. In: Portugali J (ed) *Complex Artificial Environments*, Springer, Berlin, pp 219-232
- Openshaw S (1988) Building an Automated Modelling System to Explore the Universe of Spatial Interaction Models. *Geographical Analysis* **20**: 31-46
- Oryani K, Harris B (1996) Review of Land Use Models: Theory and Application. *National Transport Library* (<http://ntl.bts.gov>), pp. 80-91
- Portugali J (2000) *Self-Organization and the City*. Springer, Heidelberg
- Prigogine I (1980) *From Being to Becoming*. Freeman, S. Francisco, CA
- Pumain D, Sanders L, Saint-Julien (1989) *Villes et auto-organization*. Economica, Paris
- Rabino GA (1993) The Blind Policy-maker and the Selfish citizen: Urban Geography and Planning in the Light of the Evolutionary Approach. Invited lecture to 8th European Theoretical and Quantitative Geographical Colloquium, Budapest (unpublished paper)
- Rabino GA (2005) *Processi Decisionali e Territorio nella Simulazione Multi-agente*. Esculapio, Bologna
- Rabino GA (2007) Dall'Informazione alla Conoscenza del Territorio. In: Sartorio G (ed) *Atti del 50° Corso di Aggiornamento in Urbanistica Tecnica* (forthcoming)
- Snow CP (1993) *The Two Cultures*. Cambridge University Press, Cambridge
- Soja E (2000) *Postmetropolis: Critical Studies of Cities and Regions*. Blackwell Publ., Malden, MA
- Torrens PM, O'Sullivan D (2001) Cellular automata and urban simulation: Where do we go from here?. *Environment and Planning B* **28**: 163-168
- Udy J (1991) *A Typology of Urban and Regional Planners*. Mellon Press, New York
- Wegener M (1994) Operational Urban Models: State of the Art. *Journal of the American Planning Association* **60**: 1: 17-30
- Wilson AJ (1974) *Urban and Regional Models in Geography and Planning*. Wiley and Sons, Chichester
- Wolfram S (2002) *A New Kind of Science*. Wooden Book, Champaign, IL
- White R, Engelen G, Uljee J (1997) The Use of Constrained Cellular Automata for High-resolution Modelling of Urban Land-use Dynamics. *Environment and Planning B* **24**: 323-343.

The Multi-Agent Simulation of the Economic and Spatial Dynamics of a Poli-Nucleated Urban Area

Ferdinando Semboloni

Center for the Study of Complex Dynamics, University of Florence, Italy

Abstract

A multi-agents simulation model for the development of a poli-nucleated urban area is presented. This model, CityDev, is based on agents, goods and markets. Each agent (family, industrial firm, commercial firm, service firm, or developer) produces goods (labor, buildings, consumption goods) by using other goods and exchanges the goods in the markets. Each agent needs a building where to live or work, hence the urban fabric is produced and transformed as the result of the co-evolution of the economic and spatial systems. The model is applied to Florence (Italy) and its main feature – the interactivity via Internet is shown. In fact web users can direct during the simulation the agents generated by the simulator as well as the new agents established by themselves. In conclusion the basic characteristics of a multi-agents method are highlighted: the comprehensive character of an agent based simulation, the ability to interact with human users, and the validation based both on observed data and on a direct interaction with real actors.

1 Introduction

Urban dynamics modeling has received a strong impulse in recent years by the application of micro-simulation methods such as cellular automata (Torrens and O'Sullivan 2001). The modeling of the spatial dynamics in urban as well as in ecological systems seems to be one of the best candidates for the application of this method. However the tools to simulate spatio-temporal systems were already developed in the field of urban-spatial analysis (Tobler 1970). Hence the cellular automata method, especially in applied research, was soon brought back to the traditional approach which in essence is based in the spatial causation, i.e. in the simulation of effects caused in the land use of the central cell, by the land uses located in its neighbor. When a temporal lag is included in this spatial causation, the model becomes a dynamical, and possibly non-linear of the reaction-diffusion type (Batty 2003). The parameters included in the model are tuned by using the method of statistical induction: the observed reality is supposed to be able to teach which are the elementary relations whose final outcome shows the best fitting with data (Straatman et al. 2004).

This evolution of micro urban analysis towards ready-to-use models which has been very rapidly sketched, is explained by the need to apply these models to the real urban context in order for them to be handled in the decision process. Nevertheless, the utilized method contrasts, partially at least, with the essence of the complexity paradigm based on the knowledge of the elementary relations and on the non-linear deduction of the emergent pattern through their reiterated application. This contrast chiefly depends on the wrong identification of the atomic element of the study, i.e. the land use with its ability to influence the surrounding cells.

In fact in urban systems what one observes is the result of the human activity which builds and transforms the existing buildings or roads thus changing the land use, or the human activity itself: people working, living, or moving. In essence what is modeled by the previous cited models is the changing of the product of human activity, i.e. the land use, and not the subject which produces these changes. In other words it is likely that the subject A produces the object O_1 , while the subject B influenced by the behavior of subject A produces the object O_2 , and the theory wants to explain the production of the object O_2 as directly depending on the object O_1 without considering the decisions of the subjects A and B . The difficulty to identify the elementary relations does result from having established objects as the atomic elements of the model.

In essence the traditional method immediately relating the observed “behavior” of objects shortens the cognitive effort, and shows its advantages in many contexts, depending of the goal of the research. But it is still not able to recognize the true atomic element of the urban dynamics: the human actor. This is why we start from the basic dichotomy: subject-object, where the subject or agent, as the basic engine, produces, transforms, and consumes the objects which are conceived as passive entities. Nevertheless, these objects, especially these belonging to the real estate, change the environment of the subjects thus influencing their decisions.

When the subject is established as atomic element also the role of space changes. While in the model based on spatial causation, space is a basic attribute of the atomic element (think for instance to the cell whose coordinates are included in its static attributes, such as the surface or the elevation), in the agent model space is not the first concern of the interaction which is independent on it.

This “subject-object” approach is related to the multi-agent method widely applied in social (Epstein and Axtell 1996) as well as in economic analysis with ACE (agents computational economy)(Tsfatsion 2002). Urban analysis as well has been recently adopted this method in connection to cellular automata (Benenson and Torrens 2004, a review is in Parker et al. 2003). Two are the main fields of urban analysis in which the multi-agent approach has been applied: that of location and that of movements. The latter has been modeled as a simulation of moving agents: car drivers (Nagel and Schreckenberg 1992), and pedestrian (Schelhorn 1999), the first has considered the most important sectors of the urban life: social segregation (Portugali and Benenson 1995; Benenson 1998) mainly based on Schelling’s segregation model, economic location especially of retail activities (Otter et al. 2001; Arentze et al. 2001), auction processes in housing market including phenomena such as gentrification (O’Sullivan 2002), and the planning

process as a bargaining among social and political actors (Arentze and Timmermans 2003; Ligtenberga 2004).

These models, unlike other large-scale urban models such as ILUTE (Miller et al. 2004), and PUMA (Ettema et al. 2005), usually focus on a sub-system of the spatial and urban analysis and establish as parameter what in the other sub-systems is a variable. Take for instance the housing market, and consider the family income as a parameter of the auction process. Of course when the labor market and the production process are considered the family income becomes a variable, and the model dynamics has to be changed if one wants the different markets to reach the equilibrium simultaneously. Now the aims of the proposed model, as ambitious as it may look like, is to establish a comprehensive simulation of the urban economic dynamics, including goods, labor, land and buildings markets, in order to evaluate the cause-effect relations crossing the border of the sub-systems, such as for instance the effects of an economic crisis on traffic, or the relation between the increase of consumption goods and building rent.

In essence the proposed model wants to simulate a whole economic system and arises from the interaction of sub-models. The behavior of consumers in a consumption goods market and in private service market gives rise to the various concentration of retailers and a private services. On the other hand, in the land market, the model for the selling of land plots is responsible for the urban shape, and the different bids of families or of firms for buildings, generate the land use pattern. The spatial pattern is highly influenced by road networks which are utilized for calculating distances. Finally, because movement of agents are calculated in the model, a minimum path model with congestion generates the traffic flows which in turn affect on the transportation cost. In this sense the proposed model is a system of sub-systems. For this reason it is able to calculate all the variables of the economic process and to make the city growing from a simple initial seed.

In addition the complete utilization of the multi-agent method makes it easy to describe the interaction with human users, which are allowed, through Internet, to manage agents by using their own strategies instead of those provided by the model for the computer agents. This feature, which is already experimented in multi-agent models (Duffy 2001), can be utilized both for comparing the performance of human and computer agents, and for planning purposes as explained later.

In conclusion, the human actor which produces and consumes products is the root of the ontology of the model. Thereby the dynamics includes subjects, i.e. agents, with cognitive and productive capabilities and objects which are passive entities but able to influence by their very existence the behavior of subjects. Markets are the virtual places in which objects are first collected and then sold to the consumers. Hence the three main classes utilized in the model are: *Agent*, *Good* and *Market*.

2 The Model

The model here presented and derived from the previous assumptions is named CityDev (see Semboloni et al. 2002; Semboloni 2005 for a further documentation). It is an urban multi-agent simulation which includes the whole aspects of the urban system and is conceived for the interaction with human users. The model focuses on the city as an economic system. Thereby the urban dynamics is simulated in its real and monetary aspects. In essence, the spatial aspect comes out from the transportation costs and from the need for each agent to utilize a building.

CityDev is based on agents, goods and markets. What connects these classes with a network of relations is the set of actions agents are allowed to perform, among which buying, production and selling are the most important. In fact, each agent produces goods by using other goods and trades the produced goods in the markets. Since in order to perform their tasks agents need a building, they are located in the physical space. Thereby, the evolution of the urban fabric results from the agents' interaction.

A grid of 100x100 squared land cells, is the spatial pattern of the simulation. The side of each land cell depends on the extension of the surface of the simulation. In the experiment shown in this paper, concerning Florence and its surrounding area, the extension of the area is 20x20 kilometers, thus the side of each land cell is 200 meters. The buildings of the city are considered as composed of indivisible 3-D cells (Semboloni 2000), whose surface is equal to that of squared land cells. Moreover, these 3-D cells can be superposed in case of a multi-floors building, thus generating a mix of land uses and a varying of the density.

The surface of land cells plays an important role in the model because in each 3-D cell only one agent is allowed to live or work. Thereby the agent does not represent a person, but a group of persons established in relation to the surface of the 3-D cell. For instance the number N of persons belonging to a family is $N = s^2/100$, where s is the side of a squared cell in meters. In other words, 100 squared meters is considered as the gross quantity of floorspace assigned to each person, and including open space, such as courtyards and gardens, secondary roads and squares. The direct and exclusive link between an agent and a 3-D building cell makes the logic of the program more clear and the computational work faster, but undoubtedly is a limitation, because we suppose an homogeneous behavior of all the persons included in the group. On the other hand this aspect has to be considered in relation to the number of persons included in each group, as well as to the global number of persons.

The dynamics of the economic urban system is demand driven such as in the economic base theory: an outside demand of industrial product stimulates the production of industry. Families are generated in order to work in industrial firms. Because families demand final consumption goods, commercial firms are generated, which in turn demand workers and so on. In addition the accelerator principle is utilized for the generation of developers in relation to the varying of the demand of buildings needed by agents for living or working.

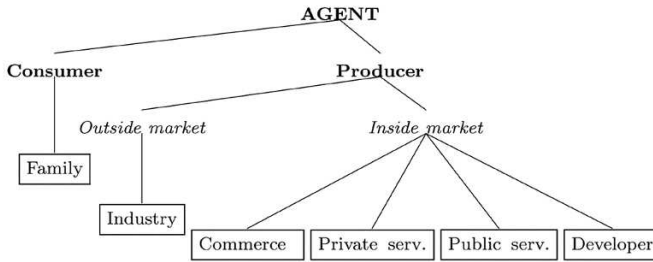


Fig. 1. The hierarchy in the agents classes. Each framed sub-class at the bottom of the hierarchy can be instantiated thus generating the individual agents.

While agents interaction is responsible for the markets equilibrium i.e. the quantity of the exchanged goods as well as their price, and for the emergence of the urban spatial pattern, the total quantity of each class of agents is managed by a meta-model which is in charge of these global aspects of the simulation. Once these global quantities are established at the beginning of each step, which is supposed to represent one year, then agents trade in the markets, selling and buying, and by using the goods bought in the markets, they finally produce their own goods. At the end of the production phase the global quantities are recalculated to allow a new step to be performed.

3 Agents Interaction and Strategies

Agents are organized in classes and sub-classes (see Fig. 1). To each class pertain specific actions which establish a relation between the *Agent* class and the other two main classes of the model: *Good* and *Market*. The general class *Agent* is allowed to perform the basic actions while the sub-classes specify these actions in relation to their role. The basic actions are: “to get”, “to produce” and “to put” (in the market). These are formalized in the following way. The agent A_i gets (i.e. buys) the good G_j which belongs to the market M_j :

$$A_i \leftarrow [G_j \in M_j]$$

The agent A_i produces the good G_j by using the goods $G_j \dots G_k$:

$$[G_j \dots G_k] \Rightarrow A_i \Rightarrow [G_i \in A_i]$$

and the agent A_i puts (i.e. sells) its good G_i in the M_i market:

$$[G_i \in A_i] \Rightarrow M_i$$

The three basic actions can be chained in order to show the complete process which is repeated for each agent at each step of the simulation: agent A_i produces

good G_i by using the goods $G_j \dots G_k$ the agent has got from markets and puts the produced good in the market, as in the following expression:

$$[G_j \in M_j \dots G_k \in M_k] \Rightarrow A_i \Rightarrow [G_i \in A_i] \Rightarrow M_i$$

Table 1. Input-output table of the relationships among agents. The *bullet* means the existence of flux: from one of the agents listed in the first column to one of the agents listed in the first row. The flux, i.e. the corresponding good or service exchanged is listed in the last column.

	Land owner	Family	Developer	Ind. firm	Commercial firm	Priv. Services firm	Publ. service	Outside demand	Product
Land owner	-	-	•	-	-	-	-	-	Land plot
Family	-	-	•	•	•	•	•	-	Labor
Developer	-	•	-	•	•	•	•	-	Building
Ind.firm	-	-	-	-	-	-	-	•	Export goods
Comm.firm	-	•	-	-	-	-	-	-	Cons. goods
Priv.services	-	•	•	•	•	-	-	-	Priv. services
Public services	-	•	-	-	-	-	-	-	Publ. services
Outside offer	-	-	•	•	•	-	-	-	Raw materials

Agents belonging to the sub-classes inherit the basic actions and specify them in relation to their environment. For instance, in case of a family, "to get" becomes: "to buy a consumption good", but also: "to get a job from the labor market". In case of an industrial firm "to get" becomes: "to buy raw material", and also: "to bid for a shed in the shed market". These actions thus result in a tree in which general actions constituting the root are specified.

The list of allowed actions stands for the agent capabilities. These have been established both using theory, especially microeconomic literature, and observation. Of course this list reveals a particular point of view which in essence is economic, and whose main aim is to sketch the urban reality in order to capture its primary features.

Because agents trade goods they have produced, a networks of relations is established among agents, as shown in Table 1. This table, is formulated as an input-output system: by row the output of each agent, and by column the input. In the last column goods or services produced or provided by agents are listed. These in-

clude: land, labor, buildings (housing, commercial, and industrial), exported goods, consumption goods, private and public services and raw materials. The organization of goods in classes and sub-classes is shown in Fig. 2.

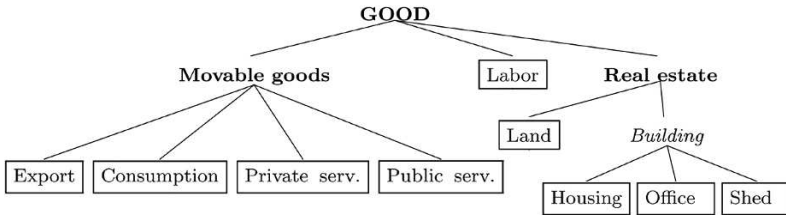


Fig. 2. The hierarchy of the main good classes. Each framed sub-class at the bottom of the hierarchy can be instantiated thus generating an individual good.

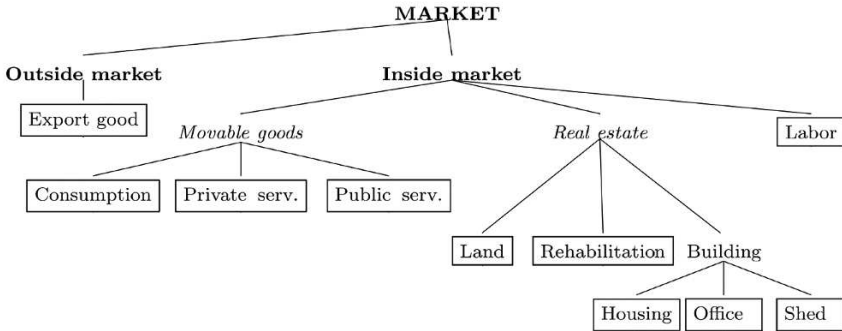


Fig. 3. The hierarchy of the main market classes. Markets cannot be instantiated, they are singleton classes. The rehabilitation market includes buildings abandoned and to be rehabilitated. In turn the building market includes building carried out by developers, or for sale because they have been abandoned by the previous renter.

To exchange goods and services, markets are provided. A specialized market exists for each good or service (in Fig. 3 the organization of markets in classes and sub-classes is shown). In each market agents offer the goods they have produced, set a price for each good and sell it to the first buyer agreeing with this price, or to the highest bidder. Two types of markets in fact exist in relation to the clearing process: that in which the consumer simply buys the desired good at the price established by the producer, and markets in which a consumer bids for the desired good which, in turn, is sold to the highest bidder. The consumption goods market and the private services market belong to the first type, while the buildings market belongs to the second one. In the buildings market (housing, commercial and industrial) a buyer first chooses the desired building and second bids for this

building. The bids for the same building are collected and the building is assigned to the best bidder.

Agent behavior is ended to the maximization of a goal. In order to maximize its goal an agent performs a set of possible actions by using strategies. These consider a set of inputs coming from the environment and establish the best choice under the existent conditions. Strategies are included in the *Strategy* class of the model, which can be accessed by the concerned agents. In fact when an agent performs an action, for instance “buy a consumption good”, its method “buy” gets the correct strategy “buy a consumption good” from the *Strategy* class. The chosen strategy takes from the agent’s method the data which represent the environment of the action (the list of the goods taken from the consumption goods market, its budget etc.) and returns the result, which in the previous case is the chosen good.

The strategies are established from the beginning, in other words there is no learning, even if the evolutionary aspect of the agent behavior is implemented through imitation, i.e. following the decision of the successful agent. Usually strategies react to a demand coming from the other agents. Pro-active strategies are utilized by agents when they plan the quantity of their production in the next step.

In order to explain how strategies have been established a brief epistemological digression is useful. The observation of the reality is a method borrowed from a natural science, namely physics. The great advantage of natural sciences in comparison with social sciences, lies in the asymmetry of the relation between the observer and the observed reality: only the observer has a cognitive capability, the only knowledge mean is observation, and the observed reality can be easily manipulated. In turn in urban studies, as in other social or economic fields the reality usually cannot be reproduced. Nevertheless, in contrast with natural context, the social system is made of entities with cognitive capabilities. While a particle does not usually interact with the observer, a developer, when inquired, may (or may not) answer and give useful information for establishing the elementary behavior of an agent in an urban multi-agent model. And this information can be utilized in the model because the entities that are formalized are similar to that observed in the reality.

For these reasons, agents’ strategies have been established by using interviews with the real actors, especially developers even if micro-economic theory and the concerned literature have been utilized as well. However the final judgment in relation to a strategy depends on the result of the simulation, both in relation to the spatial pattern as well as in relation to the equilibrium reached in the markets. Often it is useful to build a simplified model of the sectorial dynamics into which to experiment the proposed strategies. Nevertheless, when all the sectors are connected in the whole simulation the real problem is the identification of the long causal chains which produce the model result. Only intuition, and a deep knowledge of the model functioning may aid to identify the key problem and the point of the code where to make the changes which (eventually) will produce the wished effects.

Coming back to the agents, these are distinguished in two main sub-classes in relation to their main goal: consumers and producers. Families belong to the first

group, while industrial firms, commercial firms, private service firms, public services, and developers belong to the second group. Consumers maximize the quantity of consumed goods, under the constraint of an established budget, while producers maximize the earnings under the constraint of a production function. In the next section the behavior of these two sub-classes of agents is shown.

3.1 Consumers

Because in the experiment the side of the land cell is 200 meters, a family represents a group of 400 inhabitants living in the same 3-D cell, 200 of which are also workers. Families are the only final consumers. They consume goods and supply labor. Their actions are: to find a job, a house, to buy consumption goods, and to utilize public services.

Because a family has to earn for paying the rent and buying the goods, it gets a job offered in the labor market. From the job a family earns a salary and eventually a part of the profit of the firm where a family works. Under the constraint of an established budget a family spends a share (say 50 percent) of its budget for consumption goods and private services. The rest of its budgets is spent for housing and transportation costs. Consumption goods are a class including all the goods consumed by a family. In essence a family chooses the cheaper – transportation costs included – good offered in the consumption goods market and buys the maximum quantity of goods the family can afford.

In turn, in the housing market a family first chooses a house among that having a cost – rent plus transportation cost to working place – less than the established budget. Further, among the houses satisfying this condition, a family chooses that having the greatest surface, the better quality and the minimum distance from the workplace, because the action of buying a house includes the decision concerning location. The bid is calculated as the established budget minus the estimated transportation costs. However a family can change house as well as job. Roughly a family considers this eventuality with a probability proportional to the period from which the job started or the house was rented, but, before changing the house, a comparison is made between the current house and the best offer available in the market. In addition, because the rent of a house may change during the simulation, a family is constrained to look for a new house if the rent plus transportation costs overcomes the established budget.

3.2 Producers

Producers are distinguished by activity sectors: industrial, commercial and private services firms, public services and developers. The number of employed families is established and determines the size of the firm. For instance an industrial firm employs two families and its size equals an industrial firm with 400 employees. The number of families working in each class of firm is always connected with the size of the 3-D building.

Now, considering all the producers, except developers, they have common strategies related to the quantity produced, to the price of the produced goods, as well as to the location which is included in the action of buying a building where to work. The cost of production usually includes fixed costs (rent plus salaries), variable costs (raw materials) and a congestion cost proportional to the quantity produced over an established threshold. By using the marginal costs function the producer establishes the supply curve, which relates price and quantity. The number of produced goods is related to the variation of the quantity sold in the previous steps, and the price is established applying the supply curve.

In order to rent a building, and to choose the location, a producer follows the example of the existing similar producers whose activities have resulted profitable. Thereby a producer first chooses the building for which the amount of profits of similar surrounding agents is the greatest, and second bids for the chosen building an amount equal to the difference of the average sales, minus the average production cost and minus the average profit of the surrounding similar agents. As consumers, producers can change the building and the location with a probability proportional to the period from which the building has been rented, after a comparison between the advantages of the current location with those of the best buildings offered in the market.

Transport costs are paid only by industrial firms for raw materials, from the place where the firm is located to the access points to the city, i.e. a railway station, an highway tollgate, or an important external road.

We come now to the developers. They construct buildings in which families, private firms and public services live or work. Hence, they have an important role in shaping the urban fabric and are the key actors in the building process which comprises the following steps. First: the developer buys one or more land plots in the land market, choosing that with the highest difference: average price of the surrounding buildings minus the price of the land plot. Second: a developer decides which type of building to carry out in relation to the observed demand. Further, among the owned cell, chooses the best with a probability proportional to the expected bid. Third: the building is supplied in the buildings market. This buildings market has three separate sections: housing, industrial buildings (sheds), and commercial or service buildings (stores and offices). Fourth: agents bid for the desired building in the concerned buildings market and the building is assigned to the best bidder.

Because the quality of a building is decreasing during time, a building if abandoned by the renter, can be rehabilitated and offered in the market for the buildings to be rehabilitated. This happens when the difference between the potential rent and its current rent overcomes an established threshold. Developers can buy an existent building from this market, rehabilitate it and then offer it in the concerned market. This set of actions in connection with the behavior of families should be able to reproduce the both the "slumization" and the gentrification phenomena (Smith 1987).

4 Global Dynamics

While the agents interactions is governed by the allowed actions and strategies, the total quantity of agents in each class is managed by a meta-model. The urban dynamics is in fact constrained by the global system in which it is embedded. If an urban model should be able to simulate the residential location, it can say nothing about the quantity of immigrants unless its boundaries are enlarged to the other cities or regions from which the migration process is originated.

This meta-model generates firstly the growth of the outside demand which, according to the economic base theory, is the key quantity for the growth of the model because it increases the profits in the industrial sector. In fact new productive agents are generated if in the previous step the total amount of extra-profits in connection with an unsatisfied demand makes possible the existence of othersimilar firms. In turn, an agent is eliminated if his budget becomes negative, as if he has bankrupted. Consumer agents as well are submitted to a similar dynamics. In fact new families are created if the demand for workforce is greater than the existing number of families, while families with a negative budgets are canceled, as if they were emigrating elsewhere. The varying of the number of agents, in relation to the buildings stock, generates a demand of buildings which stimulates the birth of new developers. In essence the process is a mix of the demand driven economic model as supposed by the economic base theory and of the accelerator principle which relates the growth and decline of the number of developers to the demand of buildings due to the variation of the number of agents.

The process of selling of land is similarly submitted to this meta-model. In fact the number of land plots which are put up for sale is established by the relation between the demand of land cells, coming from developers, and the current offer. In addition the meta-model chooses, as in behalf of a land owner, the land plots which are put up for sale and their price. A land plot is chosen in relation to its utility for the urban development which comes up from a mix of accessibility and low density. Because the city grows at each step, the land rent is reevaluated and this additional value is transferred to the building rent. In turn the buildings quality is decreased during the simulation. Because families and firms have a limited budget the increase of rent gives rise to the relocation as well as to the upgrading of buildings.

Finally, the meta-model is in charge of assigning the commuting fluxes to the roads network and to recalculate the distance on the road network in order to include the traffic congestion in the urban dynamics.

In conclusion this meta-model acts as a general control on the agents' dynamic. This general control represents the outside conditions to the model, such as the outside demand, the land supply, and the calculation of traffic congestion and of real distances.

5 Experiments and Validation

In order to show some results, the simulation concerning the evolution of the urban area of Florence, is presented. This urban area, currently populated by about 600'000 inhabitants, is an agglomeration of cities and towns including the chief of them, Florence and other small towns which grew especially in the 80s and 90s due to the high rent of the central area and to the classical trade-off between transport cost and floorspace. The result has been a poly-nucleated area in which western part has grown faster and more widely than the eastern (see Fig. 7, left side).

It is worth noting that the needed data are limited to the DEM (Digital Elevation Model), the road network, lakes, rivers etc. The method consists in the simulation of the economic system beginning from an initial group of consumer and producer agents, and of buildings located in the center of the towns comprised in the simulation area (see Fig. 5 and 6, left side). This small seed roughly represents the state of the area in a period comprised between the medieval age and the beginning of modern development which, for Florence can be situated on 1865, when the town became, for five years, the capital of Italy.

In the further steps, the outside demand is increased in order to generate the growth of the whole system. This outside demand is a parameter which should be included in the model after an observation of the total inhabitants in each period of the simulation. Nevertheless, in order to utilize a simple method, we have chosen a logistic growth which is limited by a maximum quantity. This maximum quantity of the outside demand is calculated in relation to that currently existent in order to make the simulation realistic (see Fig. 4 in order to compare the logistic growth with that observed for the whole area in the period 1871-1991). In addition, because the area is also a central place for a large region, a part of the commercial and private service demand is supposed to come from the outside.

A general problem of a simulation such as this, aiming in reproducing the evolution of a system over a large period, is the establishment of outside conditions. In fact due to the changing economic condition two possibilities are open: to fix the rules over the whole period of simulation or take them as variable. Usually the first method has been chosen. In fact, the quantity of outside commercial demand has been considered as fixed, in relation to the total demand, even if it has grown during the period of simulation. The same method has been applied to road networks which have changed during the years. In this case only the most important urban road existing from the beginning of the simulation have been considered, excluding highways which have a limited impact on urban shape. Similarly, unbuilt areas, such as for instance the historical gardens, the area occupied by historical villas, or by the airport have been considered from the beginning of the simulation as unbuildable land. In turn, the second method has been utilized for the establishment of the points of access for industrial raw materials. Because these points have changed dramatically, from railway stations to highway tollgates, they have been considered as variable during the simulation.

The outcomes of the model concern the number of agents, the number of consumption goods and buildings supplied and bought at each period, and their ave-

rage price, the number of rehabilitated buildings etc., as well as the evolution of each agent: earnings, expenses, goods produced etc. In addition, the growth of the city can be observed as an increasing stock of buildings located somewhere in the urban area. In other words the simulation of the urban development, which is shown in Figs. 5, 6 and 7, is probably the most interesting results, but it is a very limited part of the whole outcome of the model.

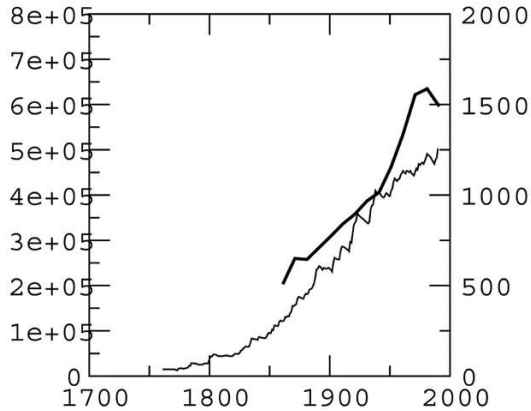


Fig. 4. A comparison of the real a simulated inhabitant. The steps of the simulation have been shifted as if the begin in 1760. Y axis: time steps; X axis: the number of inhabitants (left side), the number of agents (right side). Thick line represents the observed inhabitants, thin line the simulated inhabitants.

The results, in relation to the land use, are shown in Figs. 5 and 6. In Fig. 7 a comparison is made between the simulated pattern and a sketch of the current urban pattern in the area. This comparison is actually the most important criterion of validation of the model with the observed pattern. Other aspects which are usually considered during the simulation experiment, are the equilibrium of the average price reached in the markets, the differences between the average price of the commercial buildings, housing and sheds, the location of industrial areas, the correct concentration of the retail agents in the commercial centers and sub-centers, and the distribution of the rent of the buildings in order to test its decreasing from the center of the main town (see Fig. 8).

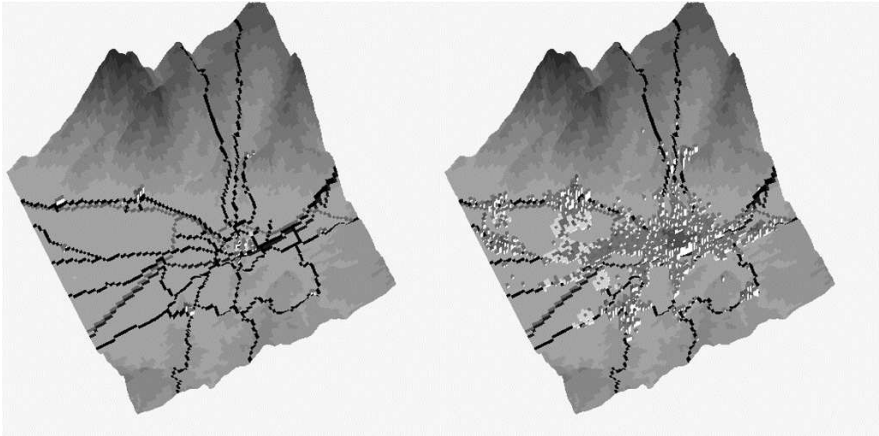


Fig. 5. The evolution of Florence and its surrounding area. Left: 20 steps, right: 230 steps.

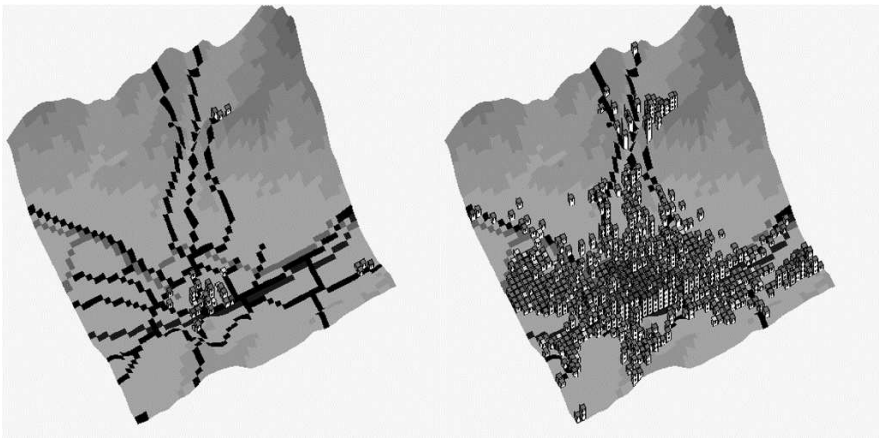


Fig. 6. The evolution of Florence. Left: 20 steps, right: 230 steps.

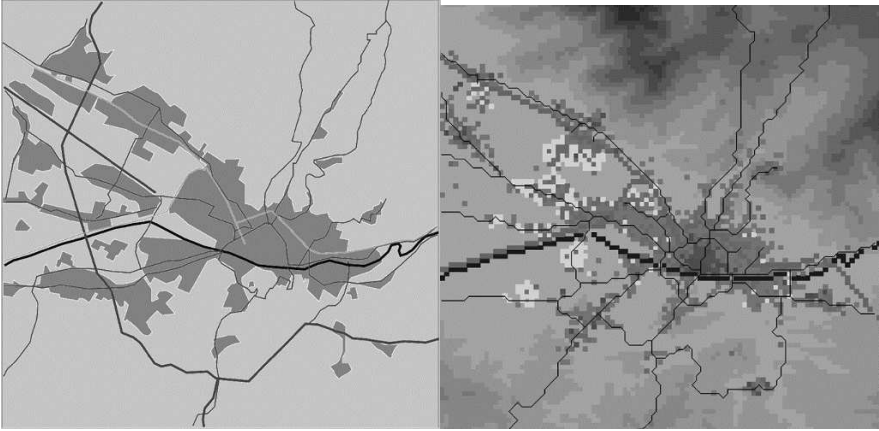


Fig. 7. A comparison with the current urban pattern. Left: a sketch of the current urban pattern. Right: the simulation at step 230.

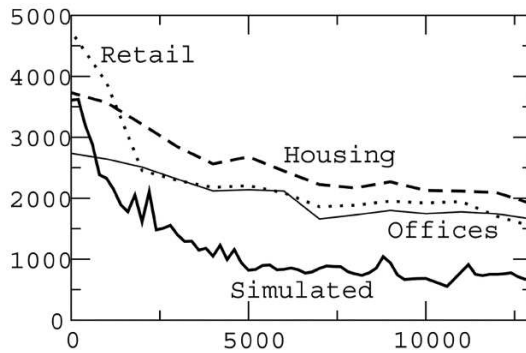


Fig. 8. A comparison of the observed and simulated building rent per squared meters of floorspace. In the X axis is the distance in meters from the city center of Florence, in the Y axis the building rent, observed in relation to retail, housing and offices, and the simulated average.

6 Calibration of Parameters

Following the method of natural science, parameters are usually included in urban models, which must be calibrated through the utilization of a lot of data, by using statistical methods. This is one of the reasons for the famous criticism of large scale models (Lee 1973). Consider for instance the weights of a neural network (Li and Yeh 2002). They do not arise from a direct observation of a reality but

only from a calibration of the model, because the parameters utilized pertain only to the model and not to the reality. In fact a method of calibration has to optimize some function of the fitness of the model results in relation to the observed data, by using a method: from the least square, in the simplest case, to the simulated annealing in the worst case.

In the CityDev model this problem is less dramatic, and the most part of the parameters can easily be observed in the reality, without having to be calibrated with the classical approach previously cited, because in the multi-agent approach the entities that are formalized are the same we observe in the reality. Take for instance the average transportation cost, the salary of a family, or the percentage of money spent for housing by a family, all these values have a meaning and pertain not only to the model, but also to the reality, from which they can easily be extracted without a calibration process.

A minor part of parameters does not correspond to some data easily observable in the reality. These are for instance parameters regulating the land cells which are put up for sale, which are responsible for the urban patterns. In this case a trial and error method has been manually applied, by experimenting the proposed parameters in a simplified model of the sectoral dynamics in order to find the best values.

7 The Interaction with Human Users and the Utilization of the Model

The model simulation is continuously under the control of human users. User interaction relies on a web interface (<http://fs.urba.arch.unifi.it:8080/suncity>), which is an essential part of the simulation. The web interface includes a home page through which web users can register or login.

The main parameters such as the growth of the outside demand and the number of the steps of the simulation can be controlled by an human user with administrative privileges. This control signifies the utilization of the model for forecasting purposes. The classical tools of the urban planning are also available for human users, such as the limitation of density, and the establishment of allowed and forbidden land uses in each land cell, as well as the creation of a new road.

In addition a human user may directly interact with the model. The interaction between human user and the model is a consequence of the multi-agent approach. In fact because the model is based on agents which replicate human actors, interaction becomes easy. Not only in the classic way, previously cited, in which a parameter is changed in order to simulate scenarios, but as a direct interaction in which a human user creates an agent, manages it, thereby interacting with the other agents managed by the model. In fact, when an user enters the simulation, he/she may choose the agents to control and consequently perform the actions allowed for these agents by using its own strategy. For instance in case of developers a human user is allowed to decide which land plot to buy, and among the owned land plots, he/she chooses where and which type of building (residential, commercial, or industrial) to build. In order to help a human user to take decisions

for the managed agents, information is available which concerns the development of the current simulation and includes various maps at different scale, along with graphs displaying the dynamics of the main variables (number of agents, buildings, prices etc.).

This interactive aspect can easily be utilized as support to the decision not only of public authority as usual, but also to the private investment decision. In addition it shows similarities with the gaming simulation practiced in the 70s (Taylor 1971), the first approach to a multi-agents simulation, but not with commercial products such as SimCity which, for what it is possible to know, looks like a land-use transport model, subject to the variation of general parameters.

8 Conclusion

Beginning from the identification of the atomic element of the urban dynamics: the human actor, we have built an ontology of the urban economic system founded on the basic dichotomy: subject (i.e. agents) and objects, i.e. goods and buildings. And because goods have to be exchanged, markets have been included as a class of the model in addition to agent and goods.

We have demonstrated how these classes are connected through the agents actions which are performed in order to maximize a goal. This goal is the maximization of the consumed goods in case the agent is a consumer, or of the earnings in case the agent is a producer. Strategies are the “intelligent” tools agents utilize in order to maximize these goals.

By using these simple assumptions a multi-agent model has been developed as a comprehensive simulation of an urban economic system, in which all the economic variables both in real and monetary aspects are managed, and the building of the urban fabric is only one of the many facets of the process.

Finally it has been shown how the validation and calibration of the model may follow a method which is partially different from that borrowed from natural science and based both on observed data and on a direct interaction with real actors.

In conclusion the multi-agent approach allows to design a subject-centered model, able to cross the border of sectorial approaches, while permitting an interaction between human users and the model dynamics which is the base for the utilization of the simulation. The simulation of the whole economic process is the key to understand the complexity of the urban dynamics, and to cross the borders of sectorial approaches.

Acknowledgment

The CityDev project has been previously funded by the University of Florence under a program on the innovation in teaching. In this first phase the project has

been developed in collaboration with Jurgen Assfalg, Saverio Armeni, Roberto Gianassi, and Francesco Marsoni (Faculty of Engineering of the University of Florence). In a second phase the project has been funded by the Italian Ministry of the University and Research. Valerio Melani (Faculty of Engineering of the University of Florence) has re-organized and further developed the CityDev model.

References

- Arentze T, Borgers A, Timmermans H (2000) A knowledge-based system for developing retail location strategies. *Computers, Environment and Urban Systems* **24**: 489-508
- Arentze T, Timmermans H (2003) A multiagent model of negotiation processes between multiple actors in urban developments: a framework for and results of numerical experiments. *Environment and Planning B: Planning and Design* **30**: 391-410
- Batty M (2003) Agents, cells and cities: New representational models for simulating multi-scale urban dynamics. CASA, Working Paper nr. 65, University College London, UK
- Benenson I (1998) Multi-agent simulations of residential dynamics in the city. *Computers, Environments and Urban Systems* **22**: 25-42
- Benenson I, Torrens PM (2004) *Geosimulation: Automata-based modeling of urban phenomena*. Wiley, New York
- Duffy J (2001) Learning to speculate: Experiments with artificial and real agents. *Journal of Economic Dynamics and Control* **25**: 295-319
- Epstein ME, Axtell R (1996) *Growing Artificial Societies - Social Science from the Bottom Up*. MIT Press, Cambridge, MA
- Ettema D, de Jong K, Timmermans H, Bakema A (2005) PUMA: multi agent modelling of urban systems. In 45th Congress of the European Regional Science Association, Vrije Universiteit, Amsterdam
- Lee DB (1973) Requiem for large scale urban models. *Journal of the American Institute of Planners* **39**: 3: 163-178
- Li X, Yeh AGO (2002) Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science* **16**: 323-343
- Ligtenberga A, Wachowicza M, Bregta AK, Kettenisb AB DL (2004) A design and application of a multi-agent system for simulation of multi-actor spatial planning. *Journal of Environmental Management* **72**: 43-55
- Miller EJ, Hunt JD, Abraham JE, Salvini PA (2004) Microsimulating urban systems. *Computers, Environment and Urban Systems* **28**: 9-44
- Nagel K, Schreckenberg M (1992) Cellular automaton model for freeway traffic. *J. Physique I* **2**: 2221
- O'Sullivan D (2002) Toward micro-scale spatial modeling of gentrification. *Journal of Geographical Systems* **4**: 251-274
- Otter HS, van der Veen A, de Vriend HJ (2001) ABLOoM: Location behaviour, spatial patterns, and agent-based modelling. *Journal of Artificial Societies and Social Simulation* **4**

- Parker DC, Manson SM, Janssen MA (2003) Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers* **93**: 314-337
- Portugali J, Benenson I (1995) Artificial planning experience by means of an heuristic cell-space model: simulating international migration in the urban process. *Environment and Planning A* **27**: 1647-1665
- Schelhorn T, O'Sullivan D, Haklay M, Thurstain-Goodwin M (1999) Strests: An agent based pedestrian model. CASA, Working Paper nr. 9, University College London, UK
- Semboloni F (2000) The dynamic of an urban cellular automata model in a 3-d spatial pattern. In: XXI National Conference Aisre: Regional and Urban Growth in a Global Market, Palermo
- Semboloni F (2005) The CityDev project: an interactive multi-agents urban model on the web. In: Portugali J (ed) *Complex Artificial Environments*. Springer, Berlin, pp 155-164
- Semboloni F, Assfalg J, Armeni S, Gianassi R, Marsoni F (2004) CityDev, an interactive multi-agents model on the web. *Computers environment and urban system* **4**: 45-64
- Smith N (1987) Gentrification and the rent-gap. *Annals of the Association of American Geographers* **77**: 462-465
- Straatman B, White R, Engelen G (2004) Towards an automatic calibration procedure for constrained cellular automata. *Computers, Environment and Urban Systems* **28**: 149-170
- Taylor JL (1971) *Instructional Planning System: a Gaming Simulation Approach to Urban Problems*. The University Press, Cambridge
- Tesfatsion L (2002) Agent-based computational economics: Growing economies from the bottom up. *Artificial Life* **8**: 55-82
- Tobler WR (1970) A computer movie simulation urban growth in the Detroit region. *Economic Geography* **46**: 234-240
- Torrens PM, O'Sullivan D (2001) Cellular automata and urban simulation: where do we go from here? *Environment and Planning B:Planning and Design* **28**:163-168

Modelling the Micro-Dynamics of Urban Systems with Continuum Valued Cellular Automata¹

Alberto Vancheri¹, Paolo Giordano¹, Denise Andrey¹, Sergio Albeverio²

¹ Accademia di architettura, University of Italian Switzerland, Mendrisio, Switzerland

² Institut für Angewandte Mathematik, University Bonn, Germany

Abstract

We present a mathematical model for urban systems based on a continuous valued cellular automaton. In the modelling we have an urban system, described through a specification cell by cell of built volumes and surfaces for different land uses and a system of agents interacting with the urban system and governed by fuzzy decision processes depending on the configuration of the urban system. For developers e.g. a point in the decision space specifies the cell and a set of continuous parameters describing the building quantitatively (e.g. surface and volume). The use of a continuum state space enables one to write a system of differential equations for the time evolution of the CA and thus to study the system from a dynamical systems theory perspective. Computer simulations on an artificial case with detailed real characteristics are presented.

1 Introduction

In the last few years many problems have been raised that require new modelling strategies about urban problems. The first one concerns the meaningful calibration of the many parameters appearing in the most developed and powerful CA models, as for example that of White and Engelen (White and Engelen 1997, 2000; Bäck et al. 1996). The problem is connected with the possibility to compare simulated configurations of the city with the empirical ones, and thus it is also connected with the important problem of models' validation. The use of numerical variables, like in our setting, instead of categorical maps enables one to apply fuzzy logic to a set of suitable chosen quantitative indicators to base the comparison of maps (see Vancheri et al. 2005; Giordano et al. 2004b).

Another problem is connected with the expected presence of bifurcations and phase transitions in these types of complex systems. These are highly typical phenomena that have been extensively studied in the field of dynamical systems, es-

¹ Research supported in part by the FNSRS 2100-060732, and the Marie Curie ERG MERG-CT-2005-014906.

pecially those described by differential equations (see Weidlich 2000; Torrens 2001; Wilson 1981; Haag 1986, Holyst et al. 2000 and references therein).

The dependence of the observed time evolution of the urban system on the initial configuration and the response of the system to an external perturbation are other examples of problems that are hard to be faced in an effective, scientifically clear and powerful way in the conceptual framework of CA. In fact such problems are mainly discussed in the framework of dynamical systems described by differential equations. To try to construct a CA and to prove that it verifies some kind of ordinary differential equation, one must at least have a quantitative description of the configuration of the CA through numerical variables. Furthermore one must be able to take a limit for the time step going to zero.

Indeed the possibility to describe the dynamics of a microscopically detailed CA model of urban dynamics using a differential equation opens the possibility to face problems about exogenous control of the system (e.g. through law indexes used in urban planning) using the mathematical theory of control for continuous dynamical systems.

A large amount of work in these type of problems has been done in the last 30 years in the field of synergetics and socio-dynamics (see Haken 1978; Haag 1989; Haag and Weidlich 1984; Weidlich 1991, 1997 and references therein).

1.1 Features of Our Model and Basic Hypotheses

In this paper we propose a generalization of the notion of CA for the study of urban growth processes with the aim of providing a suitable conceptual and computational frame taking into account the considerations above.

The main features of the model are:

- The state of a cell is not given by a *qualitative* label chosen in a finite set of possibilities, as in the more usual CA models, but by a *continuum* set of real numbers (see Sect. 2.2).
- The model contains a set of *stochastic evolution rules* containing explicitly the duration of the time step; unlike many other urban models, the randomness has been introduced with a clear urban meaning (see Sect. 3).
- Each urban process has *its own type of neighbourhood*, improving in this way the faithfulness and the flexibility of the modelling approach (see Sect. 2.5).
- In case of a Markovian dynamics (that is in case we are considering urban transformations without memory), in the limit for $\Delta t \rightarrow 0$ we *prove* a system of *ordinary differential equations* (ODE) for the mean values of the dynamical variables. More generally, even in case of a non-Markovian dynamics, we derive instead a system of *random differential equations* (RDE), which uses the concept of forward mean derivative (Nelson 1985), for the state variables; the possibility to describe memory effects using some type of differential equation is a new result (these results cannot be presented here for reasons of space, see Vancheri et al. (2004c) for an intuitive proof of the ODE, Vancheri et al.

(2004a) for a rigorous proof of the ODE and Giordano et al. (2004a) for a proof of the RDE).

It is important to note that we never suppose some kind of differential equation in the construction of the model, which is performed as in usual CA for urban systems keeping in mind a clear urban meaning; after this we can prove that necessarily the dynamics of the state variables verify a suitable differential equation.

One of the main aim of this type of work is thus to join the powerful approach of CA models, in particular its simple and detailed description of the urban system and of the forces acting in it, with the advantages given by the mathematical and conceptual tools developed in the field of synergetics and socio-dynamics, first of all the possibility to describe the system using differential (or random differential) equations.

The second aim (the principal one for this article) is an attempt to construct the model formalizing the real behaviour of the urban (stochastic) processes through the analysis of agents' decisions and their formalization using fuzzy logic methods. In spite of this it is important to immediately precise that our model is not a multi-agents model: we only foreseen the statistic behaviour of populations of agents without following the single agent. Note the conceptual difference between the description of the probabilistic laws for populations of agents (where randomness is due to the differences between the behaviour of single agents) and the description of the probabilistic laws for a single agent (where randomness is due to psychological effects, interaction with the external world, unknown effects, etc.).

In spite of this difference we can say that our type of models are able to faithfully describe the micro-dynamics of urban agents' behaviour, very similar to that obtainable with multi-agents models, giving at the same time, a very good link with mathematical theories and without the well known computational problems of these type of models. On the other hand all the state quantities are aggregate inside the cell and inside each population of agents, unlike multi-agents models where each agents hold its individuality, so that we can say that our models describe the system at a mesoscopic scale.

Moreover the continuous state space of our generalized CA permits to obtain *quantitative* outputs from computer simulations, unlike classical CA and similarly to continuous dynamical systems modelling. Hence the problem of calibration of the model can also be faced in a more effective way, e.g. using the fuzzy theory of control or meaningful quantitative indicators.

In other words we have the usual simplicity of the CA approach in the construction of a model, through the definition of simple but meaningful local rules, the microscopic details of multi-agents models, and the advantages of having differential equations which we prove for the dynamic of the state variables.

The very general hypotheses (applicable to other complex systems) our construction is based on are the following:

- We consider a set of *interacting elements*: people changing house, urban builders, public administration, CA urban cells, a whole city, etc. We will call them *agents* (note that cells of our CA will also be agents). What agents one have to chose it depends on the list of interactions one want to describe into the model.

- *Interacting these agents change their status*, which is described by continuous suitable state variables (generally speaking continuous valued or discrete).
- Every interaction is of type “agents a_1, \dots, a_n have an interaction of type α ; among them $s_1, \dots, s_c, c \leq n$ change their status”.
- Every interaction α depends by the status of the agents belonging to a suitable set, named *neighbourhood* of α (see 2.5 for examples).
- The systems is *updated at discrete time step Δt* . Inside time steps the system does not have evolution (“it is frozen”). This is why we always have to think Δt small.
- The system has a *stochastic evolution* representing our impossibility to exactly foreseen the behaviour of agents’ interactions. Using fuzzy set theory methods, agents’ decision processes can be formalized and hence randomness can be introduced in a meaningful way.
- Every *interaction produces a certain amount of goods* starting from suitable resources. Thanks to these goods the states of agents change.

As it will be more clear in the following, this new type of CA can be used as a modelling framework for several complex systems; it can also be seen as a method on the one hand to create a detailed model and on the other hand to derive for these systems differential equations not so easy to discover directly.

2 Urban Structure of The Cellular Automaton

In this section we will describe how to chose the urban structure of the CA: borders, cells, state space, type of interactions between agents, neighbourhood of an interaction and we will present the evolution rules. Each one of these elements has a clear urban meaning that contributes to faithfully describe the urban dynamics with a mathematical language.

2.1 Borders and Cells

The situation in which it is most simple to handle borders for our type of CA is the one given by an urban region with natural fixed borders, like rivers, woods, motorways, or surrounded by a large, free terrain available for urban growth; here “large” means “with respect to the maximum growth we are able to forecast in the time interval we are considering”.

The second step in constructing our model consists in a cellular decomposition of the urban space. The cellular space will be indicated with Γ . *A cell may have an “arbitrary shape”*. In some urban models it is natural to choose the shape of the cells taking into account the subdivision of the urban space determined by urban plans or by some feature of the land associated with the structure of the settlement or with the topography of the land (e.g railway, motorways, rivers, lakes). The

only condition is that inside a cell it is possible to suppose homogenous urban laws (e.g. the same indexes of the urban plan).

As in usual CA models the time is discrete. As mentioned above, the quantities relevant for the dynamics of the system contain explicitly the duration $\Delta t > 0$ of the temporal step as a parameter. In this way we can study the behaviour of the model with respect to different concrete choices of Δt and we can study the “speed of changing” of these quantities as $\Delta t \rightarrow 0$, obtaining in the limit a differential equation. We will see later that explicit criteria for the choice of the time step Δt are available (about this problem see also Bäck et al. 1996).

2.2 Continuous State Space

The state of a cell in usual CA models for urban studies is given by a qualitative label, e.g. describing the prevalent land use in the cell. As explained in the introduction, for our aims we prefer to use continuous valued variables, precisely a set of numbers giving information about *how* different land uses occupy the space inside the cell. In our model we have chosen the following quantities to define the state of a cell (note that they are all extensive quantities):

- The amount of *built* volume for residences and offices (indicated with the common label A), for small commercial surfaces (CI), for big commercial surfaces (CII), and for industries (I). We will indicate these quantities respectively by V^A , V^{CI} , V^{CII} and V^I .
- The volumes V_u^R , V_u^{CI} , V_u^O *effectively in use* respectively for residences, commerce CI and offices (the difference $V^X - V_u^X$ being the unsold or unused volume of the use X).
- The surfaces S^{Ag} used for agriculture.
- The surface S covered by buildings (we do not distinguish for this variable the different kinds of buildings).
- The surface resource Δ already employed and hence no more available for further development.
- The surface Δ^A used for residential buildings (including private gardens, parking, and infrastructure).

All these quantities are expressed by real numbers (relative to chosen units of surface and volume).

Let us give some remarks about the dynamical variables:

- We do not distinguish here built volume for residences and offices because we suppose that these land uses can be located in the same kind of space.
- We will assume that the commercial surfaces of type I are first built and then sold or given for rent. They can be occupied and abandoned following a dynamic depending on the level of the demand and rents. Commercial surfaces of kind II are built and then directly used by big commercial companies and, as a consequence, follow different dynamics than those of type I.

- The quantity $S_F = S_T - \Delta$, where S_T is the total surface of the cell computed without considering the fixed areas (e.g. water, railways), gives the amount of surface at disposal in the cell for further development.
- Empirical data about the urban region have to be chosen having in mind these dynamical variables. In Switzerland these data are available for example at the Land Registry Office or in the Office about taxation of real estate.

Grouping together the eleven dynamical quantities listed above, we obtain the state vector $v(c) \in \mathbb{R}^d$, (in this case $d = 11$) associated to the cell c :

$$v(c) = (V^A, V^{CI}, V^{CII}, V^I, V_u^R, V_u^{CI}, V_u^O, S^{Ag}, S, \Delta^A, \Delta) \tag{1}$$

When we need to explicitly indicate the time dependence of the state vector we will use the notation $v(c, t)$.

2.3 Local Parameters

The description of a cell needs to be completed specifying a set of quantities that do not have an autonomous dynamics. Some of them can vary exogenously. We will call them local parameters and we will group them in a second vector $w(c)$ associated to the cell c .

The most important quantities belonging to $w(c)$ and considered in the model are the following:

- the total distance (we use the symbol $:=$ for “equal by definition”)

$$d(v) := \sum_{v' \in V} d(v, v') \tag{2}$$

of the vertex v of the local transport network which is nearest to the cell c . More precisely we model the network as a graph, constituted by a finite set of vertices V connected by edges E . The vertices of the graph are the access point to the network. The edges are the connection lines. Given two vertices $v, w \in V$ we define the distance $d(v, w)$ as the minimum number of edges needed to move from v to w . This distance must be distinguished from the length of the shortest path between v and w , because it is defined counting edges. In the case of a subway, for example, this distance is interpreted as the number of railway stops needed to move from a station to another;

- the distance of the cell from the nearest vertex of the local transport network;
- the distances from the railway, the motorway, the motorway exits and the river, to take into account several effects as the disturbance due to the noise and the availability of infrastructure for the transportation of the goods;
- the edification index I_M^c and the covering index I_M^c given by the urban plan:

$$I_M^c := \frac{V^M}{S_T}, \quad I_M^c := \frac{S^M}{S_T} \tag{3}$$

where $S_T > 0$ denotes the difference between the total surface of the cell and the surface occupied by fixed elements as rivers, railways and streets; S^M is the maximal surface that can be covered by buildings in the cell and V^M the maximal volume that can be built in the cell;

- the surface occupied by secondary urbanizations works (e.g. green parks, parking places, schools, sport centres, theatres, administration offices);
- the surface occupied by fixed elements as rivers, railways, streets and other infrastructures.

Depending on what urban quantities we are interested in, we can put in the vector $w(c)$ besides these principal local parameters also topographical information, e.g. the slope of the terrain.

The subdivision of the relevant quantities for the descriptions of the cells into dynamical variables and local parameters is related to the idea underlying the possible use of this kind of model. Some of the local parameters associated to a cell can be modified acting on the infrastructural system of the city, some others through changes in the urban plans. In all these cases we do not model the dynamics of the related processes, but assume it as exogenous. This is equivalent to consider the parameters contained in the vector $w(c)$ as control parameters of a stochastic dynamical system. So the question is about how the system responds to modifications of the control parameters that are produced exogenously. This point of view is strongly related to the problem of the interplay between self-organization and central control of the urban system. Indeed all the processes changing the components of the vector state $v(c)$ are endogenous and are produced by decision processes of a set of agents (population, business, and companies). On the contrary the processes changing the vector of the control parameters $w(c)$ are related to centralized decision coming from the administration.

2.4 Interactions (Urban Processes)

In order to define rules for our CA, we start specifying a set A of processes that can modify the state vector $v(c)$ of a cell c . For our model we have chosen the following interactions as elements of A :

1. Occupation of free, already built, residential volume for residential use.
2. Occupation of free, already built, commercial volume of kind I.
3. Occupation of free, already built residential volume for use as an office.
4. Abandoning of residential volume.
5. Abandoning of commercial volume of kind I.
6. Abandoning of offices.
7. Construction of a residential building on a free terrain.
8. Construction of a commercial building of kind II on a free terrain.
9. Construction of a factory on a free terrain.
10. Transformation of agriculture in free terrain.

An interaction (event) of the kind $\alpha \in A = \{1, \dots, 10\}$ will be also called an α -interaction (α -event).

2.5 Neighbourhood of an Interaction

Unlike classical CA, we find more useful, generally speaking, to abandon the strictly geometrical nature of the notion of neighbourhood and to choose directly it for each type of interaction answering the following question: «*what are the cells important for this type of interaction?*». So, e.g., for the first type of process “*occupation of free, already built, residential volume for residential use*” one has to take into account all the cells near the considered one (e.g. those which are completely inside a circle of radius 200 metres), but also every single cell containing a big commercial centre not too far from the cell (e.g. less than 5 kilometres). Although for the interaction “*construction of a residential building on a free terrain*” we considered all the cells inside a circle of radius 200 metres only, for “*construction of a commercial building of kind II on a free terrain*” we necessarily have to consider a bigger neighbourhood (e.g. 5 kilometres of radius). By these examples it is clear that this simple changing of the point of view in the construction of the neighbourhoods can conduct toward a better description of the urban system’s dynamics.

2.6 Goods and Resources of an Interaction

To specify exhaustively a transformation $\alpha \in A$ we have to give $n(\alpha) \in \mathbb{N}$ continuous parameters $\pi = (\pi_1, \dots, \pi_{n(\alpha)}) \in \mathbb{R}^{n(\alpha)}$ that describe the process α quantitatively. We will call these parameters the goods produced by the transformation α . We will explain this point in some details only for the process 7 (the other cases can be described in a similar way). A rough description of a new residential building in the frame of our model requires the specification of its volume \bar{V} , of the covered surface \bar{S} , of the total surface $\bar{\Delta}$ associated to the building and of the internal subdivision of spaces, that is the volume \bar{V}^A allocated for residences or offices and the commercial volume CI: $\bar{V}^{CI} := \bar{V} - \bar{V}^A$. The surface $\bar{\Delta}$ takes into account not only the covered surface but all the space occupied by the building, including gardens, private parking and little streets for the access. It is a very important quantity because it enables to get some information about the average typology of the buildings located in the cell and it enables to evaluate correctly the amount of surface resource consumed by the building. In most cases $\bar{\Delta}$ can be identified with the surface of the parcel of terrain where the building has been built. We group together these parameters to form a vector $\pi = (\bar{V}, \bar{S}, \bar{\Delta}, \bar{V}^A) \in \mathbb{R}^{n(\alpha)}$, with $\alpha = 7$ and $n(\alpha) = 7$. In general each process $\alpha \in A$ can be seen as the production of one or more goods, through the employment of other spatial extensive quantities: the resources. For example the creation of a new volume \bar{V} causes

a reduction, at the next time step $t \rightarrow t + \Delta t$, of the volume resource $I_M^c(c) S_T - V(c,t)$ (see Eq. 3) available in the cell in accordance with the urban plan; the same is true for the good S and the surface resource resource $I_M^c(c) S_T - S(c,t)$. As a further example consider the occupation processes 1 – 3; in this case the resource is represented by the available volume in the cell for the considered use.

The goods π obviously have a stochastic nature, due to the random behaviour of the agents, whose probability distribution will be introduced in Sect. 3.2. We will call good and resources, in general, all the spatial quantities that are respectively produced or consumed in a given interaction. These concepts are thus relative to a given interaction: a good produced in an interaction can be the resource employed in another one.

2.7 Evolution of the State

The effect of an event of the kind $\alpha = 7$ on the state vector $v(c)$ of the cell c at the time step $t \rightarrow t + \Delta t$ can be described in the following way:

$$\left\{ \begin{array}{ll} t & \rightarrow t + \Delta t \\ v_1(c) = V^A & \rightarrow V^A + \bar{V}^A \\ v_2(c) = V^{Cl} & \rightarrow V^{Cl} + (\bar{V} - \bar{V}^A) \\ v_9(c) = S & \rightarrow S + \bar{S} \\ v_{10}(c) = \Delta^A & \rightarrow \Delta^A + \bar{\Delta} \\ v_{11}(c) = \Delta & \rightarrow \Delta + \bar{\Delta} \end{array} \right. \quad (4)$$

all the other components of the state vector remain unchanged. From Eq. 4 we can see that the vector γ_7 , depending on $\pi = (\bar{V}, \bar{S}, \bar{\Delta}, \bar{V}^A)$ and defined by:

$$\gamma_7(\bar{V}, \bar{S}, \bar{\Delta}, \bar{V}^A) := (\bar{V}^A, \bar{V} - \bar{V}^A, 0, \dots, 0, \bar{S}, \bar{\Delta}, \bar{\Delta}) \in \mathbb{R}^d \quad (5)$$

gives the variation of the state vector caused by the transformation $\alpha = 7$. Similar expressions can be constructed for all other processes. In general we call $\gamma_\alpha(\pi_1, \dots, \pi_{n(\alpha)}) \in \mathbb{R}^{n(\alpha)}$ the variation of the state vector of the cell caused by the transformation α . The complete evolution rules are obtained as superimpositions of elementary rules like (4).

We will illustrate this point following the example of the variable 11, the total surface Δ occupied by the different land uses in a cell c . One starts considering all the interactions that affect this variable. From the list of the elementary interactions A in Sect. 2.4 we see that the relevant processes are those given by $\alpha \in \{7, 8, 9\}$. These interactions increase the occupied surface Δ creating new buildings. This motivates the following evolution rule for the 11th component of the state vector

$$v_{11}(c, t + \Delta t) = v_{11}(c, t) + \sum_{\alpha=7}^9 \gamma_{\alpha}(\pi_1^{\alpha}, \dots, \pi_{n(\alpha)}^{\alpha}) \tag{6}$$

In this formula obviously one suppose that all the three interactions 7, 8 and 9 happen in the considered time step. Generally speaking for a given dynamical variable some interactions will increase the variable and some others will decrease it.

Note that e.g. $\gamma_{\alpha}(\pi_1^{\alpha}, \dots, \pi_{n(\alpha)}^{\alpha})$ depends on Δt and hence the previous equality permits to glimpse the possibility to obtain a differential equation for $v_{11}(c, t)$ as $\Delta t \rightarrow 0$: the essential point in the proof will be to factor the terms $\sum_{\alpha=7}^9 \gamma_{\alpha}(\pi_1^{\alpha}, \dots, \pi_{n(\alpha)}^{\alpha})$ as $\Delta t X(t, v(c, t))$ for $\Delta t \rightarrow 0$ (see Vancheri et al. 2004a, 2004b and Giordano et al. 2004a).

3 Stochastic Structure of the Cellular Automaton

In this section we introduce the probability distributions which regulate the application of rules like (4) and hence the stochastic dynamics of the CA. Let us consider an interaction $\alpha \in A$; as we sketched above, a basic idea of the model is, intuitively speaking, that the more the state of the neighbourhoods of a cell c is in favour of this transformation α , the higher are the intensities of the α -events that will happen at the next step in the cell c . More precisely if the state of the neighbourhoods favours an α -interaction, then we will have a high rate of production of goods associated to α . From this point of view it is natural to ask that this “number of requests” is a Poisson process.

3.1 Counting the Number of Events: Poisson Processes

The next element that must be defined in order to complete the description of the rules for the stochastic CA is the probability density $p_i^{\alpha}(\pi, n)$ that n events of the kind $\alpha \in A$ and with goods $\pi \in \mathbb{R}^{n(\alpha)}$ happen in the cell during the time interval $[t, t + \Delta t)$. In general these probability densities depend on the configuration of the system in the neighbourhoods associated to the interaction α and on some global variable depending on the state of the system, as we will see in Sect. 4 (this is why the probability density $p_i^{\alpha}(\pi, n)$ depends on time t).

The quantity $P_{n,B}^{\alpha} := P_{n,B}^{\alpha}(t) := \int_B p_i^{\alpha}(\pi, n) \cdot d\pi$ give thus the probability that n events of type α take place with goods in the (measurable) subset B of the space $\mathbb{R}^{n(\alpha)}$.

The possibilities of choice for the probability densities $p_i^\alpha(\pi, n)$ can be strongly restricted analyzing more deeply the nature of the urban interactions which happen in a given cell c .

Indeed, let us consider a given cell c and a short time interval $[t, t+\Delta t)$. It is possible that, during this interval, nothing happens in c or, on the contrary, that we have one or more events, each characterized by some discrete label $\alpha \in A$ and a set of goods $\pi \in \mathbb{R}^{n(\alpha)}$. An appropriate question about the assignment of the probability densities $p_i^\alpha(\pi, n)$ for the transitions rules is the following:

Starting from the current state of the automaton, what is the probability $p_{n,B}^\alpha$ to have n events of the kind $\alpha \in A$ and goods $\pi \in B \subseteq \mathbb{R}^{n(\alpha)}$ in the cell c during the time Δt ?

It is reasonable to assume that the probabilities $p_{n,B}^\alpha$ are given by a Poisson distribution (we are indeed about in the same situation that we find when we want to model the process counting the number of calls received in a call-centre or the number of customers that arrive in a shop during a given time (see Kingman (1993) for basic notions about Poisson processes). Let $N_{\alpha,B}(c, t)$ be the stochastic variable that counts the number of events of the kind α that happen in the cell c during the time interval $[t, t+\Delta t)$, with continuous parameters $\pi \in B \subseteq \mathbb{R}^{n(\alpha)}$ (in the following we will also use the simplified notation $N_{\alpha,B}(c)$ if the time dependence is clear from the context).

We recall that the main condition for the use of a Poisson distribution is the independence of the events that are counted during the time Δt . We can assume that this condition is fulfilled if Δt is rather short, so that the following conditions are true:

- during the time interval $[t, t+\Delta t)$ the information about the decision of the agents does not spread out in the system, so that we can consider *independent* the number of events which happen in disjoint time intervals contained in $[t, t+\Delta t)$;
- the number of events that happen during the time Δt is small enough so that the change of the state of the system does not significantly affect the strategy of the agents. In little more precise terms: the “intensity (number of α -events per unit of time)” is constant during $[t, t+\Delta t)$. Moreover this intensity is finite;
- the fact that there will happen n events in $[t_1, t_2) \subseteq [t, t+\Delta t)$ depends only on n and on the length $t_2 - t_1$;
- for a very small Δt there will happen only either 0 or 1 event.

With these assumptions, suitably formalized, it is possible to *prove* that we necessarily have the following probability distributions for the counting variables (so that taking a Poisson distribution is not actually an assumption, but a logical consequence of the above mentioned meaningful assumptions):

$$P[N_{\alpha,B}(c,t) = n] = \frac{1}{n!} \exp(-\lambda_{\alpha,B}(c,t)\Delta t) \cdot (\lambda_{\alpha,B}(c,t)\Delta t)^n \tag{7}$$

where the positive-valued functions $\lambda_{\alpha,B}(c,t) > 0$, depending on the cell c through the configurations of the system and hence on the time t , are the intensities of the Poisson process. Note that the intensities are constant during each step $[t, t+\Delta t)$ (indeed we have a Poisson process during this time interval only, not for a greater time) and change at the end of each step because the neighbourhood state changes. Roughly speaking $\lambda_{\alpha,B}(c,t)$ is the velocity with which the events happen in time, indeed it is measured in “number of events per unit of time”. Note also that we are using here an innocuous abuse of language with respect to the mathematical theory of Poisson processes (Kingman 1993) where the intensity is the quantity $\lambda_{\alpha,B}(c,t)\Delta t$ and where usually $\Delta t = 1$.

Concretely we will introduce (see Sect. 4) the intensities $\lambda_{\alpha,B}(c,t)$ using a density

$$\lambda_{\alpha,B}(c,t) = \int_B \lambda^\alpha(c,\pi,t) \cdot d\pi \tag{8}$$

The conditions stated above give us relevant, quasi-empirical criteria for the choice of Δt . Generally speaking the actual choice of Δt depends on the size of the system, and the stated conditions are usually fulfilled if this quantity is not greater than a few days. Because Δt is short it is possible to replace the Poisson distribution with binomial, multinomial or Bernoulli distributions.

3.2 Probability Distributions for the Goods of an Urban Interaction

When a (α,c,π) -event take place in a cell $c \in \Gamma$ a value $\pi \in V^\alpha$ of the produced goods must be randomly extracted in the space V^α of the goods associated to the α -process. This can be done by means of a probability density $\beta^\alpha(c,t)$ defined in V^α . This probability density is associated to the conditional probability that an α -event has associated goods π given it took place into the cell c . The probability density $\beta^\alpha(c,t)$ is hence characterized by the relation:

$$\lambda^\alpha(c,\pi) = \lambda^\alpha(c) \cdot \beta^\alpha(c,\pi) \tag{9}$$

where $\lambda^\alpha(c,\pi) = \lambda^\alpha(c)\beta^\alpha(c,\pi)$ is the total intensity for α -events in c .

In practical cases it is much more easier to define the total intensity $\lambda^\alpha(c)$ and the probability density $\beta^\alpha(c,\pi)$ separately and subsequently the density of intensity $\lambda^\alpha(c,\pi)$ for (α,c,π) events through the Eq. 9. We will return on this point in the Sect. 4.3.

4 Fuzzy Decision Theory for Urban Processes

4.1 Decision Dynamics

In this section we will explain at some extent how the intensities of the processes (Eq. 8) are derived in our model from the dynamics of population of agents.

We will consider a set P^k , $k = 1, 2, \dots, K$ of population of agents, each made of N^k individuals. Individuals in the same population are identical, in the sense that they cannot be distinguished from the point of view of the criteria describing their behaviour. In this sense agents in this model have not an individual identity as in multi-agent model. Nevertheless agents have cognitive proprieties driving their behaviour, as we will see in the next subsection.

We will describe a decision process of an agent $a \in P^k$ as a sequence of more elementary events. First the agent, who initially is in a “passive state”, becomes active with respect to a given kind of elementary process $\alpha \in A$, in the sense that it begins a decision process which could eventually lead to an (α, c) -event in some cell $c \in \Gamma$. After an agent has become active we assume that during its sojourn in the cell c an active agent processes information concerning c in order to chose one among the following two possibilities: 1) to perform an (α, c, π) -event for some $\pi \in V^\alpha$, (after this event he will return in a passive state in the same or in another population); 2) to abandon the decision process becoming a passive agent again.

In order to follow this approach to the dynamics of agents we have to define the following quantities:

- The probability $p_A^\alpha(c, t, \Delta t) = \lambda_A^\alpha(c, t) \Delta t + o(\Delta t)$ that a passive agent in the population k starts an α -decision process during a time step $[t, t+\Delta t)$ in a cell $c \in \Gamma$ (here A means “activation”).
- The probability $p_L^\alpha(c, t, \Delta t) = \lambda_L^\alpha(c, t) \Delta t + o(\Delta t)$ that an active agent in the cell c abandon the decision process during a time step $[t, t+\Delta t)$ (here L means “Leave the decision process”).
- The probability $p_0^\alpha(c, t, \Delta t) = \lambda_0^\alpha(c, t) \Delta t + o(\Delta t)$ that an active agent in the cell c perform an (α, c, π) decision for some $\pi \in V^\alpha$ during a time step $[t, t+\Delta t)$.

In the previous list we have introduced the intensities of the processes (that is the probabilities for unit time) through the first order Taylor expansion $p(c, t, \Delta t) := \lambda(c, t) \Delta t + o(\Delta t)$ valid for a short time step Δt . We recall that the intensity of a process gives the expected number of events during the time Δt .

In order to describe in some more details the dynamics of the active agents we will do now some particular assumptions; it will become clear in the next part of the exposition that these assumptions could be relaxed and modified in a plenty of ways. The assumptions made in this example is mainly motivated by a sake of simplicity.

Let us consider as an example a generic process α involving agents belonging to a population \mathbf{P}^E made of N^E individuals (here E means “example”).

First we will assume that the activation intensity $\lambda_A^\alpha(c, t)$ for an agent in the population \mathbf{P}^E is proportional to a quantity $\Lambda_A^\alpha(c)$ determined by global, state dependent, factors only; this quantity can thus be thought as a global activation intensity. Further we will assume that the probability that a globally activated agent focuses its attention to a cell c is proportional to a local attractiveness $F^\alpha(c)$ of the cell taking values in $[0, 1]$ (we will give later a motivation of the choice of this range). These assumptions are summarized in the following expression, showing that the total activation intensity is distributed among the cells proportionally to the local attractiveness $F^\alpha(c)$.

$$\lambda_A^\alpha(c, t) = \Lambda_A^\alpha(t) \cdot \frac{F^\alpha(c, t)}{\sum_{c' \in \Gamma} F^\alpha(c', t)}. \quad (10)$$

Second we will assume that the rate of abandoning of the research process does not depend neither on the cell nor on the state of the system: $\lambda_L^\alpha(c, t) = \Lambda_L^\alpha$, for each $c \in \Gamma$. This means that an active agent abandons the decision process in average in a time $t_L^\alpha = 1/\Lambda_L^\alpha$ (if it does not perform an α -event before abandoning).

The last assumption concerns the intensity $\lambda_0^\alpha(c, t)$ of (α, c) -events, which will be assumed proportional to another function $G^\alpha(c, t)$ measuring the local attractiveness of the cell c and taking values in $[0, 1]$:

$$\lambda_0^\alpha(c, t) = \Lambda_0^\alpha \cdot G^\alpha(c, t).$$

The most important difference between the measure of attractiveness $F^\alpha(c)$ and $G^\alpha(c, t)$ is given by the different kind of decisions they are aimed to describe and by the related different point of views about the cell. The function $F^\alpha(c)$ is proportional to the probability that an active agent chose the cell c among all the cells contained in the set Γ in order to pass to the more detailed phase of decision described by $G^\alpha(c, t)$. This last decision is not a choice among many possibilities but rather a yes/not decision concerning only the cell c . Hence it is reasonable to assume that the information concerning the cell c contained in $F^\alpha(c, t)$ is rougher than the one contained in $G^\alpha(c, t)$ and corresponds to slightly different cognitive approaches to the evaluation of the cell c . We will return to this point in the next subsection.

4.2 Fuzzy Decision Theory

In this section we will sketch without getting into many details the method used to construct the quantities $F^\alpha(c, t)$ and $G^\alpha(c, t)$ used to model the dynamics of active

agents. For both quantities we make use of a fuzzy decision theory based approach.

Let us consider a decision maker who has to choose one among several possibilities belonging to a decision space A . The task consists in attaining a set of goals G_1, \dots, G_n without violating a set of constraints C_1, \dots, C_n . The specific feature of the approach based on fuzzy logic is to treat goals and constraints as fuzzy sets on the universe A : for each $x \in A$ the value $\mu_{G_k}(x)$ of the membership function associated to the goal G_j is a measure of the degree of attainment of the goal. In the same way the membership functions μ_{C_j} associated to the constraints, measures at what extent the constraint are fulfilled. A fuzzy set D is constructed on A intersecting fuzzy sets associated to goals and constraints:

$$D = G_1 \cap \dots \cap G_n \cap C_1 \cap \dots \cap C_m$$

The intersection of sets is the counterpart in set theory of the logical connective AND; thus the decision set D expresses the simultaneous attainment of the goals and fulfilment of the constraints. The decision maker chooses the element $x^* \in A$ which maximize the membership function μ_D of D . Let us notice that in fuzzy logic approach to decision theory goals and constraints are treated in the same way and hence we will not use this distinction any more.

In fuzzy logic there are many available operation, called t-norms, on the membership functions of fuzzy sets that correspond to intersection. Furthermore many methods are available in order to weight goals and constraints. It is also possible to use different kind of operations than intersections to take into account situations where goals can compensate each other or where the attainment of a critical number of goals is sufficient to have a good evaluation of a possibility.

In our model of urban dynamics the decision space consists in general in a subset $A \subseteq \Gamma$ of the cellular space. Goals and constraints of α -agents involved in an α -process are specified in terms of indicators of the urban configuration: an indicator $I_j^\alpha(c,t)$ is associated to each goal G_j^α . If for instance the goal is expressed in natural language with the sentence:

“Availability of commercial volumes around the cell” the relevant indicator could be the commercial volumes for inhabitant in a suitable neighbourhood $U(c)$ of the cell:

$$I(c,t) = \frac{\sum_{c' \in U(c)} V^C(c',t)}{\sum_{c' \in U(c)} n^R(c',t)} \tag{12}$$

where $V^C(c',t)$ and $n^R(c',t)$ are respectively the commercial volume and the number of resident population in the cell c . A membership function μ_j^α acting on the indicator $I_j^\alpha(c,t)$ describes the degree of attainment of the goal G_j^α .

In many cases an indicator can be itself the membership function of a fuzzy set on Γ . Many activities like e.g. commerce take advantage from a central position in the city because of a more intense flow of pedestrians. In order to describe goals connected with a seek of centrality we have constructed an indicator that take into account some typical features of central areas like an high degree of mixing of several kind of (non industrial) activities and a high density of the settlement. The fuzzy set C of the central cells can be described through a suitably defined membership function μ_C which can be used by different kind of agents in order to express some of their goals through expression in natural language like “very central cell” or “not too much central cell”.

We will give now a brief description of the indicator of centrality μ_C . The first step consists in defining the ratio of the volumes $V_{R,C,O}$ used inside the cell respectively for residences (R), small commercial surfaces (C) and offices (O) over the total volume $V := V_R + V_C + V_O$:

$$f_{R,C,O} := \begin{cases} \frac{V_{R,C,O}}{V} & \text{if } V \neq 0 \\ 0 & \text{if } V = 0 \end{cases} \quad (13)$$

In the second step we construct fuzzy sets M_R , M_C and M_O with associated membership functions μ_R , μ_C and μ_O representing cells with a high ratio of volume occupied respectively by residences, commerce and offices. Different membership functions are needed for different uses, because of the different meaning of the words “high ratio” in the three cases. Let us consider further the fuzzy set D of cells with a high density of built volume with an associated membership function μ_D .

A weighted mean of the membership functions μ_R , μ_C and μ_O gives the membership function μ_M of the fuzzy set M of cells with a degree of mixing of the different activities:

$$\mu_M := \alpha_R \cdot \mu_R + \alpha_C \cdot \mu_C + \alpha_O \cdot \mu_O \quad (14)$$

where $0 \leq \alpha_R, \alpha_C, \alpha_O \leq 1$ and $\alpha_R + \alpha_C + \alpha_O = 1$.

The fuzzy set C of central cells is built as an intersection of fuzzy sets M and D : $C = M \cap D$. In this way central cells are characterized as those with an high degree of mixing of activities AND a high density of the settlement.

The intersections in Eq. 11 can be realized through the t-norm given by the product:

$$\mu_C(c,t) = \mu_M(c,t) \cdot \mu_D(c,t) \quad (15)$$

More in general one can start with a description of an urban typology T in a natural language and subsequently try to individuate a set of “semantic traits” S_1, \dots, S_k characterizing the typology T . Each semantic trait is characterized through an indicator $I_k(c,t)$ like e.g. Eq. 12 and a set of membership functions μ_k acting on the values of I_k and measuring the degree of presence of the corresponding trait. If this semantic traits can be compound through logical operation

like AND, OR, NOT and modifiers like “very”, “not too much”, “few” it becomes possible to use the machinery of fuzzy logic to describe the urban typology T as a fuzzy set on Γ with an associated membership function μ_T . This fuzzy set can be subsequently used in order to express decision criteria of agents.

We point out here that these indicators of typology aim to capture some aspects of the way individuals perceives and categorizes the urban space in which they move and act. Centrality indicators like μ_C and others defined along similar lines can be useful to detect features of the urban space with respect to the quality of life in cities.

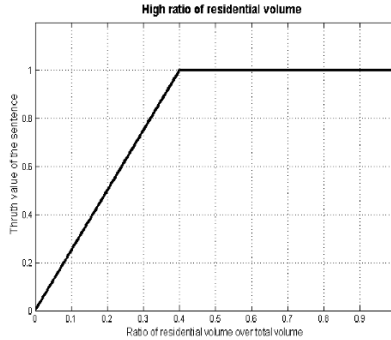


Fig. 1. Example of a fuzzy membership function

4.3 Example: Cost of Terrain, Rents and Residential Sector

We have connected the fuzzy logic based approach sketched in the previous Sect. with a method of real estate estimation worked out by the Swiss architects Nägely and Wengen (NW, see Nägely and Wenger 1997). That enable us to introduce in the model several quantities playing a crucial role in the dynamics of urban systems like rents and price of terrains. The method of NW is based on the evaluation of the relationship of a terrain with the urban and regional system where the terrain is located. They assume that the value W of an estate can be split into two contributions: the value W_E of the estate considered independently from the urban context and a positional contribution W_p . The first value can be identified with the construction costs and is roughly proportional to the volume V of the building:

$$W_E = w_E \cdot V \tag{16}$$

The authors assume that the ratio

$$q := \frac{W_p}{W_E} \tag{17}$$

of these values is a quantity depending only on the urban context concerning the terrain. They provide the user of the method with a very exhaustive check list of

prototypical situations aiming at sketching urban contexts. Following the method the user comes to an evaluation Q of the terrain in a scale ranging from 0 to 10. An empirically determined function f associates a value of the ratio q to each value of Q . The value W of the estate can be evaluated from the relations (16) and (17):

$$W = W_E + W_P = W_E \cdot (1 + q) = w_E \cdot V \cdot (1 + q). \quad (18)$$

This evaluation enable to obtain for instance an estimate of the rent per unit of volume, which is yearly a fraction γ (the capitalization rate) of the total value (18):

$$a = w_E \cdot \gamma \cdot (1 + q) \quad (19)$$

The cost of a terrain can be estimated in the following way. Let us consider a not yet built terrain for residential use with an associated edification index I_E . The method of NW can be used in order to estimate the maximum value W_{\max} that a building can have exploiting the terrain at the maximum extent allowed by the urban plan. If the surface of the terrain is Δ , the maximum exploitation corresponds to the maximum value $V_{\max} = \Delta \cdot I_E$ of the built volume. That leads to an estimated value given by:

$$W_{\max} = w_E \cdot V_{\max} \cdot (1 + q) = w_E \cdot \Delta \cdot I_E \cdot (1 + q) \quad (20)$$

It is now reasonable to assume that the cost of the terrain will be strongly influenced by the value of W_{\max} . Let us notice that the expression (20) include the dependence of the cost of a terrain both on the urban context and on the urban plan. If the model explicitly include a land market sector, a suitable chosen function of W_{\max} can be used as base price. If such a land market model is not included we can use a function of W_{\max} directly as an estimate of the cost of the terrain.

In order to use the method of NW in our model we have identified the evaluation Q normalized in the range $[0, 1]$ (a measure of attractiveness of a terrain for a given land use) with the membership function of a fuzzy set on the cellular space Γ . In order to do that we have translated in fuzzy logic terms the criteria of the check list provided by the author which is written in natural language. The functions $G^\alpha(c, t)$ introduced in the previous Sect. have been obtained from $Q^\alpha(c, t)$, the fuzzy logic version of the evaluation of NW, and the related quantities like the rent or the cost of the terrain. If for example P^{Pop} is the population of the city and α is the process consisting in the occupation of free residential volume, we can obtain from the combined method of NW the following quantities: 1) a measure $Q^\alpha(c, t)$ of the attractiveness of the cell c for an α -agent involved in an α -process; 2) the average rent $a(c, t)$ for residential use in the cell c . Now it is reasonable to assume that the active k -agent has two main goals in evaluating the cell: $G_1 \leftrightarrow$ “the cell has a high attractiveness” and $G_2 \leftrightarrow$ “the rent $a(c, t)$ is not too high” These goals can be evaluated by means of two membership functions μ_1 and μ_2 acting respectively on the indicators $G^\alpha(c, t)$ and $a(c, t)$. In this way we arrive at the following membership function for the decision set D :

$$\mu_D^\alpha(c, t) = \mu_1^\alpha(Q^\alpha(c, t)) \cdot \mu_2^\alpha(a(c, t)) \quad (21)$$

where the intersection operation has been realized through the t-norm given by the product of membership functions. The function $G^\alpha(c,t)$ is now written as a given power of the membership function in order to enhance the weight of highly attractive cells with respect to the less attractive:

$$G^\alpha(c,t) = (\mu_D^\alpha(c,t))^\beta, \quad \beta > 1. \quad (22)$$

The functions $F^\alpha(c,t)$, evaluating the attractiveness of the cell c at a less detailed level of information are written using similar method based on fuzzy logic. At this rougher level of analysis instead of considering all the information at our disposal about the cell we consider the relation of the cell with a set of distinguished points and structures on the natural as well urbanized space like rivers and lakes, airports, railway stations, motorway exits, schools, sport centres, green parks, downtown, etc. Each cell is characterized by the relations of accessibility with respect to all these distinguished points determined by the different transport networks. Adding information concerning the use of the land at a middle and large scale like the accessibility to commercial areas, the position with respect to the industrial zones, and the density of the settlement in the space around the cell we obtain a characterization of the cell taking into account its relations with the urban system. Goals and related indicators and membership functions can hence be constructed to obtain an expression like (22) for the measure of attractiveness $F^\alpha(c,t)$.

As last example let us consider the dynamics of the building sector. The use of the positional method of NW combined with methods of fuzzy decision theory enable us to evaluate the expected profit $p(c,t)$ of a constructor investing in a cell c :

$$p(c,t,V,\Delta) = \frac{C_c \cdot V \cdot (1 + q(c,t)) - C_t(c,t) \cdot \Delta - C_c \cdot V}{C_t(c,t) \cdot \Delta + C_c \cdot V} \quad (23)$$

where $C_t(c,t)$ is the cost of the terrain and C_c the construction costs. In this way a constructor can evaluate for each cell c the maximum profit that can be obtained with the constraints given by the urban plan:

$$\bar{p}(c,t) = \max_{(V,S,\Delta) \in D^\alpha(c,t)} p(c,t,V,\Delta) \quad (24)$$

where (V,S,Δ) are respectively the volume of the building, the covered surface and the surface of the terrain, α is the process of construction of new residential volume and $D^\alpha(c,t) \subseteq V^\alpha$ is the subset of the space of the goods compatible with the urban plan in c . The quantity $\bar{p}(c,t)$ defined in Eq. 24 is the indicator corresponding to the goal $G_1 \leftrightarrow$ "maximize the profit". This goal, together with other describing the risk of the investment ("how much volume should be built?") enables to construct a membership function for the decision set in the building sector.

The indicator of profit (Eq. 23) represents a typical example of the method employed to construct the probability distributions $\beta^\alpha(c,t)$ defined in Sect. 3.2 for the extraction of the goods. In this case the decision space is the space of the goods V^α associated to the process. The agents, in this example constructors, have a list of goals and constraints, the last ones given by the master plan. Oversimplifying the

example we assume that the only goal of the constructor is to maximize the profit $p(c,t,V,\Delta)$. The fuzzy set G associated to this goal and the related membership function μ_G (a function of the indicator of the profit) enable to evaluate different choices of the goods $\pi = (V,S,\Delta)$ with respect o the expected profit. The fuzzy set associated to the constraints of the master plan is in this case a crisp set C containing all the allowed values of the goods. The decision set D can be obtained multiplying the membership functions associated to the goal and the constraint:

$$\mu_D(c,t,V,S,\Delta) = \mu_G(p(c,t,V,\Delta)) \cdot \mu_C(V,S,\Delta) \tag{25}$$

The probability distribution is now obtained normalizing a power of the membership function:

$$\beta^\alpha(c,t,\pi) = \frac{(\mu_D(c,t,\pi))^p}{\int_{V^\alpha} (\mu_D(c,t,\pi))^p d\pi} \tag{26}$$

where $\pi = (V,S,\Delta)$.

5 Evolution Algorithm

In the preceding sections we have discussed all the elements which defines our model: the urban structure of the CA, its stochastic structure connected with the evolution rules and the fuzzy theory sector which permits to meaningfully connect the stochastic structure with realistic urban decision processes. If we proceed with the most natural synchronous evolution algorithm for this CA, we have firstly to evaluate the intensities of each process in each cell. Using a Poisson distributed random numbers generator we have to extract the number of α -events which happen in each cell. Then using a random numbers generator distributed as the goods produced by α -interactions, extract the goods for each transformation and finally we apply the evolution rules.

The following asynchronous evolution algorithm, based on the connection of the dynamics of our CA with the theory of Markov jump processes (MJP), is more efficient than a synchronous one because it uses less Poisson extractions.

1. Let us consider the CA at the beginning of a generic step n . Let the time be given by $t_n = (n - 1) \Delta t$, where, as before, Δt is the discrete time step. Let $N^{\alpha,X}(c,t_n)$ be the number of interactions involving α -agents with $X \in \{A,L,0\}$ (activation, abandon and decision) which happen in c during the time interval $[t_n, t_n + \Delta t)$. These variables are supposed to be independent, so that the quantity

$$N(t_n) := \sum_{\alpha \in A} \sum_{X \in \{A,L,0\}} \sum_{c \in \Gamma} N^{\alpha,X}(c,t_n) \tag{27}$$

(“number of interactions of some type in some cell”) is again Poisson distributed with intensity

$$\lambda(t_n) := \sum_{\alpha \in A} \sum_{X \in \{A, L, 0\}} \sum_{c \in \Gamma} \lambda^{\alpha, X}(c, t_n) \quad (28)$$

The formula (28) represents a sum “over every event which may happen in the system”. In (28) $\lambda^{\alpha, A}(c, t_n) := \lambda_A^\alpha(c, t_n) N_P^\alpha(c, t_n)$ and $N_P^\alpha(c, t_n)$ is the number of α -passive agents; whereas for $X \in \{L, 0\}$ we have $\lambda^{\alpha, X}(c, t_n) := \lambda_X^\alpha(c, t_n) N_A^\alpha(c, t_n)$ and $N_A^\alpha(c, t_n)$ is the number of α -active agents in the cell c .

We can thus answer the question “*when do the next interaction of the CA (of some type $\alpha \in A \cup \{A, L, 0\}$ in some cell c) happen?*” extracting an exponentially distributed random number $\theta \geq 0$ with intensity $\lambda(t_n)$. We interpret a situation in which $\theta > \Delta t$ as “the first event does not happen during the n -th step but in $k \Delta t \leq \theta < (k + 1) \Delta t, k \in \mathbb{N}, k \geq 1$ ” (note that in this case the intensities do not change up to the k -th step because the state of the automaton does not change).

2. “*Where do this first interaction happen?*”. It is possible to prove that we must answer this question using the discrete distribution

$$\left(\frac{\lambda_c}{\lambda} \right)_{c \in \Gamma}, \text{ where } \lambda_c := \sum_{\alpha \in A} \sum_{X \in \{A, L, 0\}} \lambda^{\alpha, X}(c, t_n) \quad (29)$$

At the end of this step of the algorithm we have a time θ and a cell c .

3. “*What kind of event among abandoning, activation and decision will happen?*”. It is possible to prove that we must answer this question using the discrete distribution

$$\left(\frac{\lambda_c^X}{\lambda} \right)_{X \in \{A, L, 0\}}, \text{ where } \lambda_c^X := \sum_{\alpha \in A} \lambda^{\alpha, X}(c, t_n)$$

At the end of this step of the algorithm we have a time θ , a cell c and a type of event $X \in \{A, L, 0\}$.

4. “*Of what type $\alpha \in A$ is the interaction that happen in the cell c at the time θ and corresponding to the event $X \in \{A, L, 0\}$?*”. It is possible to prove that we must answer this question using the following discrete distribution:

$$\left(\frac{\lambda^{\alpha, X}(c, t_n)}{\lambda_c^X} \right)_{\alpha \in A} \quad (30)$$

At the end of this step we have a time θ , a cell $c \in \mathbb{Z}$ and a process $\alpha \in A$.

5. If $X = A$ (activation) then $N_A^\alpha(c, t_n + \theta) = N_A^\alpha(c, t_n) + 1$ and $N_P^\alpha(t_n + \theta) = N_P^\alpha(t_n) - 1$
6. If $X = L$ (abandon) then $N_A^\alpha(c, t_n + \theta) = N_A^\alpha(c, t_n) - 1$ and $N_P^\alpha(t_n + \theta) = N_P^\alpha(t_n) + 1$
7. If $X = 0$ (decision) then we have to ask “*How much of the goods does the α -interaction produce?*”. We generate the values $\pi_1, \dots, \pi_{n(\alpha)}$ of the goods using the

probability $\beta_\alpha(c,-)$ and update the configuration of the system applying the rules (see Sect. 2.7).

8. We repeat the steps 1 – 7 of the algorithm while the condition $t \leq t_n + \Delta t$ is fulfilled. The intensities $\lambda_x^\alpha(c, t_n)$ are modified only when the whole time step is passed ($t > t_n + \Delta t$).

6 The Artificial Case Study ArtCity and Simulations

We present here simulations based on an artificial but realistic urban system called “ArtCity” (artificial city).

6.1 Description of ArtCity

ArtCity is a square grid of about 22 km x 22 km, each cell of the grid corresponds to a square of 200 m x 200 m. The system has different zones: some residential zones, a main city, agriculture, commerce, industry, a river, a motorway, a railway line (see Fig. 1 below). Note that most of the surface has no specific use so that it is free for the evolution.

The main city is placed in the centre of the urban system, so that its growing is possible in all directions. The city area is divided in 7 sub-zones: the centre and his outer belt, three intensive residential zones and two extensive residential zones. All zones have a mix of residence, commerce I and offices. The first intensive zone near the centre and his outer belt, has predominantly offices (69% of the built volume) and commerce I (20%). This zone should represent the zone of the city where commercial and financial volume is most concentrate. Furthermore the city has several public infrastructures: a town hall with a square, six schools, an opera house and a theatre, a hospital, two sport centres and a big sport area.

The residential zones are zones with less urbanization typical for a rural or a countryside zone. There are one-family houses (for example young families with children who will not live in the city) or houses for peasants, no higher than two floors. These zones are all next to the agriculture and have some commerce (a small general store or a handicraft shop) and no offices (for example a bank or a post office), except for the two zones on the left side of the urban system where there are only houses. With agriculture we do not intend the surface left when we remove from whole territory of the urban system the buildings and industry area but the surface which is really exploited for agricultural purposes. The used surface pro cell varies from 10'000 m² to 40'000 m².

The vacant rate of built residential volume in the main city is 1%, for the built office volume it is from 8% in the centre and his outer belt to 12% in the extensive residential zones. In the residential zones (the zones near to the agriculture) the vacant rate of built volume stays invariant: 0,5% for the residences, 1% for the commerce I.

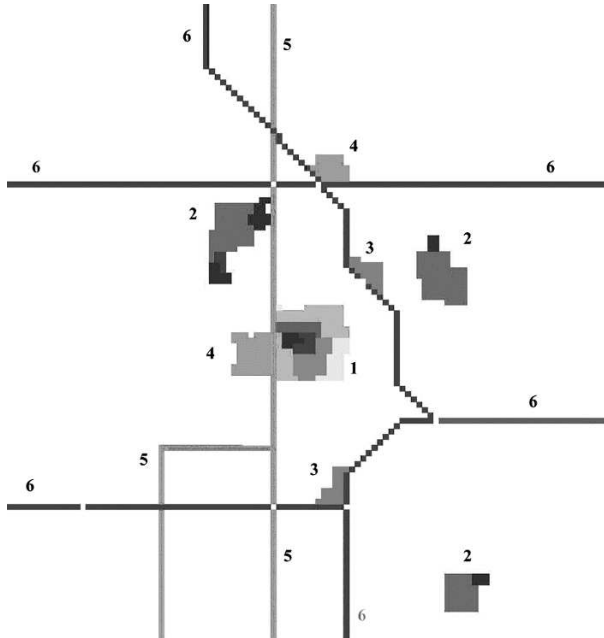


Fig. 2. ArtCity and its zones: Main city (1), Residential zones and agriculture (2), Commerce 2 (3), Industry (4), River (5), Railway and Motorway (6)

In ArtCity there are two commercial zones of type II, both near to a motorway junction. This guarantees that they are easily reachable from every point of the urban system. In the commercial zone of type II we consider only big commercial surfaces and not small shops or stores as (commerce of type I).

There are two industrial zones in ArtCity: one is opposite to the main city with respect to the river and another is far away from the city, but near to the motorway and the railway. In the first case the industry is not much noisy (for example manufacturing). We could interpret this zone as an older industry zone, where the closeness to the river is principally given by his navigability. The second zone instead could be considered as the new noisy industrial zone, it is far away from the city, but near to the modern transportation network.

The river has a width of 150 near the industrial zone, otherwise it is 120 m wide. The residual surface of the cells crossed by the river is considered as public equipped green park and can therefore be considered as a fixed land use with no evolution.

The transport network consists of a

- motorway (4 lanes with 18 m of width) that crosses the urban system in an articulated way and has 5 junctions,
- railway line that is simply a horizontally straight line about 6 km below the centre of the city and

- web of main roads (see Fig. 2). In ArtCity we do not consider the side streets. This fine web of main streets is together with the motorway very important, in particular to determine the integration of the cell (an indicator describing the relation of the cell to the whole transport network).

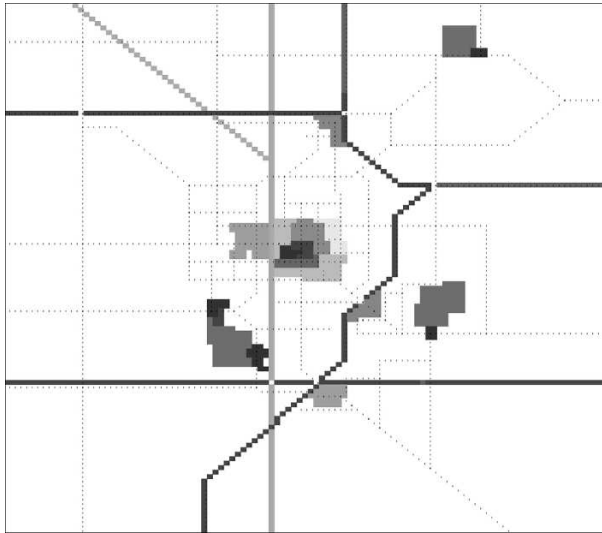


Fig. 3. The transport network of ArtCity

Since the possible interactions as described in Sect. 2.4 are creation, abandon or occupation of volumes in our case as prism, it is important to define the standard height of the building floor for each use: in ArtCity the height for residences and offices is 3 m, for commerce (type I and II) is 4 m, and for the industry it is 5 m.

Every cell of the urban system has a masterplan constraint (covering and edification index and the maximum height of the building), even if in the initial configuration there is no volume and the cell has not a specific use.

Starting from the described urban system ArtCity we did computer simulations, which will be presented in the next Sect.

6.2 Computer Simulations

The following computer simulations, describe the time evolution of two dynamical quantities, specified cell by cell:

- The total amount of built volume.
- The volume of buildings for residential use, including the part occupied by offices and small commercial surfaces.

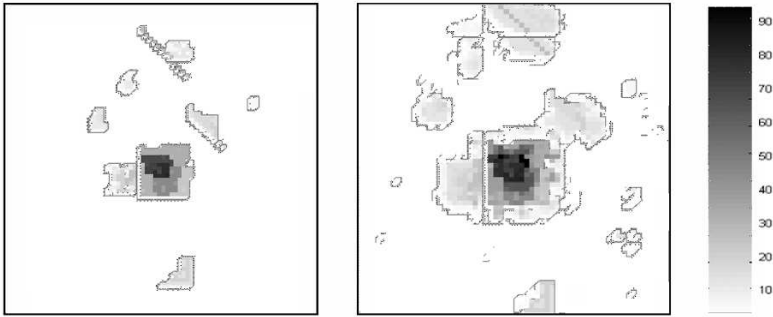


Fig. 4. Time evolution of total built volume over 20 years. Colour map ranges between 0-575

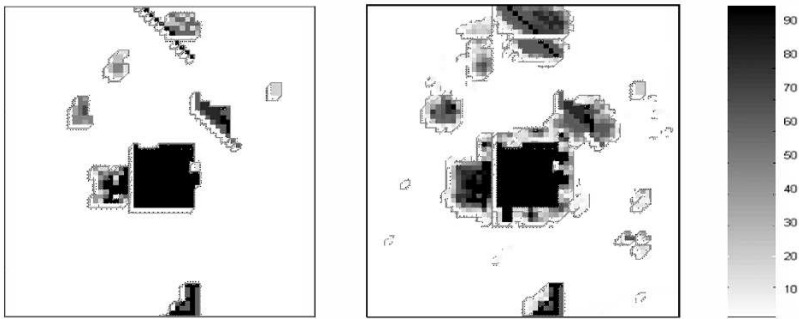


Fig. 5. Time evolution of total built volume over 20 years. Colour map ranges between 0-100

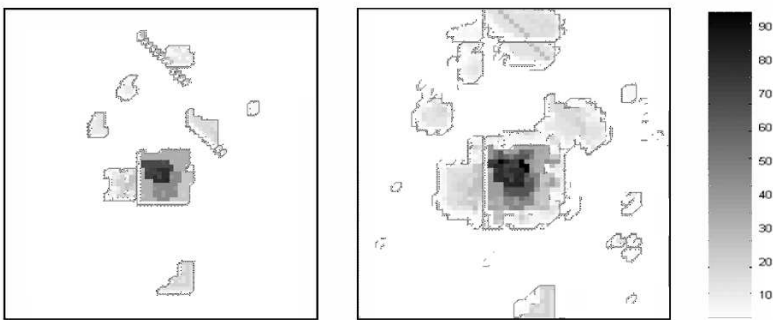


Fig. 6. Time evolution of residential volume over 20 years. Colour map ranges between 0-575

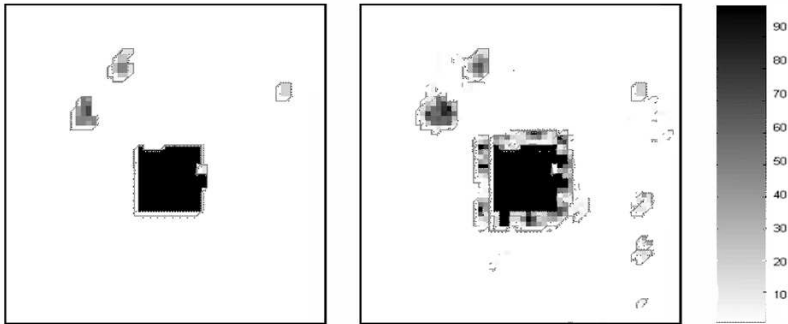


Fig. 7. Time evolution of residential volume over 20 years. Colour map ranges between 0-100

7 Conclusions

In this paper we have introduced a new type of CA for modelling urban systems. The main steps can be summarized in the following way:

- The space state of a cell is constituted by a set of continuous extensive quantities like built volumes and occupied surfaces for different land uses.
- The evolution rules are given by the probability distribution of stochastic, Poisson distributed jumps corresponding to several kinds of processes on the urban space. Each jump produces a set of continuous extensive parameters, called goods, that describes quantitatively the process. The whole stochastic sector has been introduced trying to respect a meaningful mathematical formalization of agents' urban behaviour. The use of fuzzy decision theory is a key point here.
- The typology of each cell is reconstructed using fuzzy logic methods through the state vectors in a suitable set of neighbourhoods.
- The decision process of an agent is described as a random motion in the corresponding decision space. Even if this is not a multi agents models, it tries to describe with a great detail the microscopic statistical behaviour of population of agents.

References

- Albeverio S, Andrey D, Giordano P, Simona M, Vancheri A (2003) Continuum state cellular automata based models of urban growth processes. Contributed talk at ICIAM 2003, Sydney, July 2003.

- Bäck T, Dörnemann H, Hammel U, Frankhauser P (1996) Modelling urban growth by cellular automata. In: Voigt HM, Ebeling W, Rechenberg I, Schwefel HP (eds) *Parallel Problem Solving from Nature IV Proceedings of the International Conference on Evolutionary Computation*. Springer, Berlin
- Bandemer H, Gottwald S (1995) *Fuzzy Sets, Fuzzy Logic, Fuzzy Methods. With Applications*. John Wiley & Sons, Chichester
- Giordano P, Vancheri A, Andrey D, Albeverio S (2004a) Complex urban dynamics described using random differential equations. *Accademia di architettura (USI)*, working paper, <http://www.mate.arch.unisi.ch/ACME>
- Giordano P, Vancheri A, Andrey D, Albeverio S (2004b) Reflections on the concepts of model's calibration and validation. *Accademia di architettura (USI)*, working paper, <http://www.mate.arch.unisi.ch/ACME>
- Haag G, Dendrinos DS (1983) Toward a Stochastic Theory of Location: A non Linear Migration Process. *Geographical Analysis* **15**: 269-286.
- Haag G (1986) A Stochastic Theory of Sudden Urban Growth. *Konzepte SFB 230, Heft 25*, pp 255-271
- Haag G (1989) *Dynamic Decision Theory: Application to Urban and Regional Topics*. Kluwer Academic Publishers, Dordrecht
- Haag G, Weidlich W (1984) A Stochastic Theory of Interregional Migration. *Geographical Analysis* **16**: 331-35
- Haken H (1978) *Synergetics. An Introduction. Nonequilibrium Phase transitions in Physics, Chemistry and Biology*. Springer, Berlin
- Holyst JA, Kacperski K, Schweitzer F (2000) Phase transitions in social impact models of opinion formation. *Physica A* **285**:
- Kingman JFC (1993) *Poisson processes*. Oxford University Press, Oxford.
- Langlois A, Phipps M (1997) *Automates cellulaires : application à la simulation urbaine*. Hermès, Paris
- Munz M, Weidlich W (1990a) Settlement formation. Part I: A dynamic Theory. *The Annals of regional Science* **24**: 83-106
- Munz M, Weidlich W (1990b) Settlement formation. Part II: Numerical Simulations. *The Annals of regional Science* **24**: 177-196
- Nelson E (1985) *Quantum Fluctuations*. Princeton Series in Physics, Princeton University Press
- Nägely W, Wenger H (1997) *Der Liegenschaftenschätzer (The Estate Value Estimator)*. Schulthess Polygraphischer Verlag, Zürich
- Popkov YS, Shvetsov VI, Weidlich W (1998) Settlement Formation Models with entropy operator. *The Annals of Regional Science* **32**: 267-294
- Portugali J (1999) *Self-Organization and the City*. Springer, Berlin
- Schweitzer F (ed) (1997) *Self Organization of Complex Structures – From Individual to Collective Dynamics*. Gordon and Breach Science Publisher, Amsterdam
- Schweitzer F (2002) *Brownian Agents and active particles. On the Emergence of Complex Behavior in the Natural and Social Sciences*. Springer, Berlin
- Semboni F (2000) *Automa cellulare urbano in un pattern 3-D. Studio della dinamica (Urban Cellular Automaton in a 3-D Pattern. A Study of the Dynamics)*. Proceedings of the 21st Conferenza Italiana di Scienze Regionali, Palermo
- Torrens PM (2001) Cellular automata and urban simulation: where do we go from here?. *Environment and Planning B: Planning and Design* **28**: 163-168.

- Vancheri A, Andrey D, Giordano P, Albeverio S (2005) Modelling urban dynamics with continuous valued cellular automata and fuzzy logic: a real case studies. In preparation.
- Vancheri A, Giordano P, Andrey D, Albeverio S (2004a) Ordinary differential equation for the mean value of the dynamical variables. Accademia di architettura, working paper, <http://www.mate.arch.unisi.ch/ACME>.
- Vancheri A, Giordano P, Andrey D, Albeverio S (2004b) A Model for Urban Growth Processes with Continuum State Cellular Automata and Related Differential Equations. July 2004, <http://ssrn.com/abstract=633969>
- Weidlich W, Haag G (1983) Concepts and models of a quantitative sociology. Springer, Berlin
- Weidlich W, Haag G (1987) A dynamic phase-transition model for spatial agglomeration processes. *Journal of Regional Science* **27**: 529-569
- Weidlich W (1991) Physics and Social Science – The Approach of Synergetics. *Physics Reports* **204**: 1-163.
- Weidlich W (1997) From fast to slow processes in the evolution of urban and regional settlements structures. In: Schweitzer F (ed) *Self Organization of Complex Structures – From Individual to Collective Dynamics*. Gordon and Breach Science Publisher, Amsterdam, pp 475-488
- Weidlich W (2000) *Sociodynamics – A Systematic Approach to Mathematical Modelling in the Social Sciences*. Harwood Academic Publishers, Amsterdam
- White R, Engelen G (1997) Multi-Scale Spatial Modelling of Self-Organizing Urban Systems. In: Schweitzer F (ed) *Self Organization of Complex Structures – From Individual to Collective Dynamics*. Gordon and Breach Science Publisher, Amsterdam, pp 519-535
- White R, Engelen G, Uljee I, Lavalle C, Ehrlich D (2000) Developing an Urban Land use Simulator for European Cities. In: K. Fullerton (ed) *Proceedings of the 5th EC-GIS Workshop held in Stresa, Italy, June 1999*, European Commission, Joint Research Centre, Ispra, Italy, pp 179-190
- Wilson G (1981) *Catastrophe theory and bifurcation applications to urban and regional systems*. Croom Helm, London
- Yager RR (1977) Multiple Objective Decision-making Using Fuzzy Sets. *Intl. J. Man-Machine Studies* **9**: 375-382

Multiplicative Processes and City Sizes

Damián H. Zanette

Consejo Nacional de Investigaciones Científicas y Técnicas, Centro Atómico Bariloche and Instituto Balseiro, Bariloche, Argentina

Abstract

In this contribution, I address the function of multiplicative stochastic processes in modelling the occurrence of power-law city size distributions. As an explanation of the result of Zipf's rank analysis, Simon's model is presented in a mathematically elementary way, with a thorough discussion of the involved hypotheses. Emphasis is put on the flexibility of the model, as to its possible extensions and the relaxation of some strong assumptions. I point out some open problems regarding the prediction of the detailed shape of Zipf's rank plots, which may be tackled by means of such extensions.

1 Introduction

My talk in the workshop on Dynamics of Complex Urban Systems was aimed at reviewing the role of multiplicative stochastic processes in models for city growth. In the distribution of city sizes –where “size” is identified with population, i.e. with the number of inhabitants– that role is revealed by the occurrence of Zipf's law. I have put the emphasis on the mathematical framework provided by Simon's model, and discussed a few extensions of the model to explain variations of Zipf's law –not only in the distribution of city sizes, but also in surname frequency and word abundance in human languages. My main conclusion was that, in spite of the different nature of these systems, the ubiquity of Zipf's law can be understood in terms of common underlying multiplicative processes, as captured by Simon's model.

In the discussion that followed my talk, and in subsequent discussion sessions, it became clear that the potential of Simon's model as an explanation for Zipf's law, as well as its limitations, are not well understood. This impression is reinforced as soon as one inspects current literature on the subject of city size distributions. In a recently published handbook on urban economics (Gabaix and Ioannides 2004), for instance, we read: “*Simon's model encounters some serious problems. In the limit where it can generate Zipf's law, it ... requires that the number of cities grow indefinitely, in fact as fast as the urban population.*” As I show later, not only is this assertion wrong but the truth happens to be exactly the opposite! Leaving aside the “logical” path that may have led to this false conclusion (Gabaix 1999), such strong statements are prone to become small dogmas, at

least, for the part of the scientific community which does not have the tools for critical analysis of their derivation.

Consequently, I think that it may be useful if I devote this contribution to give a pedagogical presentation of Simon's model in the frame of the evolution of city size distributions. The explicit statement of the hypotheses that define the model should already expose its limitations but, at the same time, should clarify its flexibility regarding possible generalizations. Of course, a certain degree of mathematical detail will be unavoidable. In any case, however, the emphasis will be put on a qualitative description of the basic processes involved in the modelling.

In the next section, I introduce an elementary model for the evolution of a population based on stochastic processes, and discuss the concurrent role of multiplicative and additive mechanisms in the appearance of power-law distributions. After an outline of the main features of Zipf's rank plots in the distribution of city sizes, I go on to the presentation of Simon's model in its original version, describing its implications as for the population distribution in urban systems. Then, I discuss a few extensions of the model, aimed at capturing some relevant processes not present in its original formulation. Finally, I close with a summary of the main results and some concluding remarks.

2 Multiplicative Processes in Human Populations

Human communities, as any other biological population, are subject to a multitude of actions of very disparate origins. Such actions involve a complex interplay between factors endogenous to the population, and external mechanisms related to the interaction with the ecosystem and with physical environmental factors. The fluctuating nature of those actions calls for a description based on stochastic –i.e., random– processes. Within this kind of formulation, it is explicitly assumed that the parameters that govern the evolution of the population can change with time in irregular ways. For instance, the change in the number $n(t)$ of individuals within a population during a certain time interval Δt can be modelled by means of the discrete stochastic equation

$$n(t + \Delta t) - n(t) = a(t)n(t) + f(t) \quad (1)$$

where $a(t)$ and $f(t)$ are random variables. At each time step, their values are drawn from suitably chosen probability distributions. As a consequence of the random variation of $a(t)$ and $f(t)$, the number $n(t)$ also displays stochastic evolution. Equation 1 is used to predict the statistical properties of $n(t)$, for instance, finding the probability distribution $P(n, t)$ that the population has a value n at time t . This kind of equation has been studied in detail by several authors in various contexts (Sornette 1998; 2000 and references therein).

The first term in the right-hand side of Eq. 1, $a(t)n(t)$, represents the contributions to the evolution of n that are proportional to the population itself, thus called *multiplicative*. If the population is closed, multiplicative processes are restricted to birth and death, and $a(t)$ stands for the difference between the birth and death rates

per individual in the interval Δt . In open populations, the number of individuals is also affected by migration processes. Emigration flows are generally proportional to $n(t)$ because, on the average, each individual has a certain probability of leaving the population per time unit. On the other hand, immigration has both multiplicative and *additive* effects. In fact, immigration can be favored by a large preexisting population –as in big cities– but a portion of arrivals may also occur as a consequence of individual decisions that do not take into account how large the population is. Additive contributions are described by the second term in Eq. 1. This term can also stand for negative effects on the population growth, such as catastrophic events where a substantial part of the population dies irrespectively of the value of $n(t)$ (Manrubia and Zanette 1999).

The probability distribution $P(n,t)$ of the population n , as derived from Eq. 1, can have a complicated analytical form depending on the specific distributions chosen for $a(t)$ and $f(t)$. It is nevertheless known that, for large times, it decreases as a power law,

$$P(n,t) \sim n^{-1-\gamma} \quad (2)$$

over a substantial interval of values of n . The exponent $\gamma > 0$ is given by the distribution $p(a)$ for the random variable $a(t)$, as the solution of the equation (Sornette 1998)

$$1 = \int p(a)(a+1)^\gamma da \quad (3)$$

Equation 2 holds under very general conditions on the probability distributions of $a(t)$ and $f(t)$, provided that $f(t)$ is not identically equal to zero. In other words, a power-law distribution is obtained when both multiplicative and additive processes are in action. The case $f(t) \equiv 0$ is, mathematically speaking, a singular limit. In the absence of additive processes, $P(n,t)$ becomes a log-normal distribution.

The empirical observation of a power-law distribution in a real system would require to have access to many realizations of the evolution of the same population –which, in practice, is rarely possible– or, alternatively, to follow the parallel evolution of several sub-populations of the same type. This second instance is often met in human populations, which are naturally divided into communities of different nature, determined by historical, geographical, sociocultural, and/or economic factors. One of such divisions is given, precisely, by urban settlements. In this case, $P(n,t)$ can be interpreted as the probability of having a city of size n at time t within the region that encompasses the whole population under study. In view of the above discussion, it is expected that the populations of different cities follow, under suitably homogeneous conditions over the studied region, a power-law distribution. As is well known, in fact, they do. Power laws in the population distribution of human groups of various kinds have been reported by several authors and, notably, by the philologist G. K. Zipf (Zipf 1949). The power-law dependence of the frequency of groups as a function of their population is now known as Zipf's law.

Zipf's law is often presented in an alternative formulation which, in the frame of the distribution of city sizes, goes as follows. Take all the cities under consideration, and rank them in order of decreasing population, so that rank $r = 1$ corresponds to the largest city, $r = 2$ to the second largest, and so on. Then, the population n of a city decreases with its rank as a power law,

$$n(r) \sim r^{-z} \quad (4)$$

over a wide range of values of r . The exponent z is usually referred to as the Zipf exponent. Equations 2 and 4 are closely related. In fact, the formulation of Zipf's law in terms of the probability distribution $P(n,t)$ and the rank formulation are equivalent, though the latter is much less significant than the former from the viewpoint of a statistical description. To understand the connection between the two formulations, it is first useful to recall that—in our interpretation of the stochastic growth equation 1 as describing the parallel evolution of several sub-populations—the sum

$$\sum_{n=n_1}^{n_2} P(n,t)$$

is the probability of having a city with population between n_1 and n_2 . Accordingly, the product of this sum times the total number of cities under consideration, is the number of cities with populations within that interval. Since the rank r of a city of population n equals the number of cities with populations larger than or equal to n , we have

$$r(n) = M \sum_{n'=n}^{\infty} P(n',t) dn' \quad (5)$$

where M is the total number of cities. This establishes the relation between the rank-frequency dependence and the probability distribution P . In particular, replacing Eq. 2 into Eq. 5, we get $r(n) \sim n^{-\gamma}$, implying that the Zipf exponent and γ are related as

$$z = \frac{1}{\gamma} . \quad (6)$$

This defines the connection between the power laws in Eqs. 2 and 4.

3 Zipf's Law and City Sizes

The application of Zipf's rank analysis to urban settlements implicitly assumes that individual cities are well-defined entities. Actually, however, the modern city is such a complex of intermingled systems that it defies a definition in terms of traditional classification schemes, and requires a wider concept of class (Portugali 2000). Figure 1 illustrates the fact that, while individual urban settlements can be

distinctly identified in some regions, in other places the situation is by far less obvious. Anyway, it is currently accepted that the entities to be considered in Zipf's analysis are the clusters resulting from the growth and aggregation of initially separated settlements. A plot of the population n versus the rank r for the cities of a given country or region usually reveals three regimes. For the lowest ranks, corresponding to the largest cities, the variation of n with r is generally irregular, with a marked descending step between the first one to three cities and the following. The biggest urban settlements in any large country or region often lie outside any significant statistical regularity, both within the region in question and between different regions. As the rank becomes higher, these irregularities smooth out, and the plot enters the power-law regime. In the usual representation of n versus r in log-log scales, this regime is revealed by a linear profile, typically extending from $r \approx 10$ to ranks of the order of a few to several hundreds. The Zipf exponent z , given by the slope of the linear profile in the log-log plot, is considerably uniform between different regions. It is customary to quote the value $z \approx 1$, though it may vary between 0.7 and 1.2. Finally, for the highest ranks the power-law regime is cut off, and n declines faster as r grows. Figure 2 illustrates these typical features for 276 metropolitan areas in the USA, according to data from the year 2000 census. Note carefully that the class of "metropolitan areas" does not necessarily include all urban settlements above a certain size. Below, I comment on related methodological problems in the construction of rank plots.

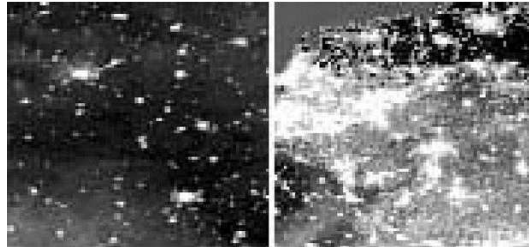


Fig. 1. Two satellite images of the Earth by night. Left: Central Ukraine. Right: North-western Germany. Each image covers an area of, roughly, 500x500 Km². Source: visibleearth.nasa.gov.

It is clear from the discussion on multiplicative stochastic processes in Section 2 that, among the three regimes identified in rank plots, the natural candidate to be explained in terms of such mechanism is the central power-law range. It also results from our discussion that, to derive a power-law city size distribution, it is necessary to take into account both multiplicative and additive contributions to the evolution of the population. These ingredients are captured by Simon's model, which is presented in next section. Let me here point out that, to explain Zipf's law in the distribution of city sizes, a model solely based on multiplicative processes –namely, Gibrat's model (Gibrat 1931) – is often invoked. As already commented, however, purely multiplicative mechanisms can only produce a log-normal distribution. While over restricted ranges a log-normal distribution may

seem to exhibit a power-law decay, $P(n,t) \sim n^{-\lambda}$ with $\lambda \approx 1$, it certainly cannot fit the variety of Zipf exponents found in real city size distributions.

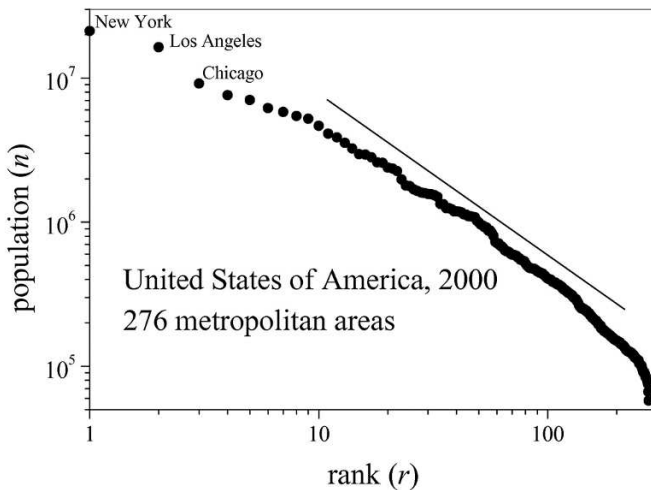


Fig. 2. Zipf rank plot for 276 metropolitan areas in the United States, after results of the census in 2000. Source: factfinder.census.gov. The straight line has slope 1.1.

Before passing to the formulation of Simon's model for the power-law regime of rank plots, it is pertinent to discuss a few aspects regarding the description of the two remaining regimes –those corresponding to the lowest and the highest ranks. As for the former, the biggest cities in a large country or region are, almost invariably, special cases that elude inclusion in any statistical description. It would be hopeless to pretend that, for instance, Paris, Berlin, or Rome enter the same statistical class as the European cities whose present population is below, say, one million. The political and economic role of those cities has been –and still is–markedly peculiar. In consequence, their individual evolution is exceptional among urban settlements and must be dealt with as such. While it would make little sense to discuss the case of the biggest cities in the frame of a statistical model for the distribution of city sizes, it is nevertheless interesting to advance that Simon's model assigns a special role to those cities: their sizes bear information on the initial state of the urban system, before the smaller settlements played any significant role in the population statistics. In this sense, Simon's model also recognizes that the biggest cities are special cases.

As for the cut-off region of highest ranks, let me mention that it is found not only in rank plots for city sizes, but also in many other instances where Zipf's law holds for intermediate ranks. A classical example occurs in the frequency of words in human languages (Montemurro and Zanette 2002; Zanette and Montemurro 2005). In the case of city sizes, the appearance of the cut-off is well known but, to my knowledge, there is no systematic study regarding the population-rank fun-

ctional dependence in that regime. This lack of quantitative empirical results discourages modelling of city sizes for high ranks, as there is no reference data to validate potential models. Moreover, as pointed out in connection with Fig. 2, the regime of high ranks is susceptible of methodological errors related to possible data incompleteness. While, arguably, the lists of large cities provided by most sources of demographic information are exhaustive, the same sources may result to be less reliable when it comes to smaller urban settlements. Inspection of many public-domain databases immediately reveals lack of completeness in the lists of cities for high ranks. The direct effect of these “gaps” is that the assigned ranks are lower than in reality, with the consequent reinforcement of the cut-off (cf. Fig. 2). Avoiding this effect without restricting too much the range of ranks under consideration, requires relying on presumably complete data sets –typically, from official census reports. This, in turn, limits the corpus of data, because such databases are not always available. In any case, as stated above, the cut-off in rank plots can be observed in other systems where this kind of methodological error is not present. In Section 5, I adapt to the case of city sizes an extension of Simon’s model put forward to give a semi-quantitative explanation of the cut-off regime for the case of word frequencies in language.

4 Formulation of Simon’s Model

Elaborating on an idea previously advanced by Willis and Yule (1922), H. A. Simon proposed the model that now bear his name in 1955 (Simon 1955), as an explanation for the origin of power-law distributions and Zipf’s law. Simon presented his model by referring to the case of word frequencies, which Zipf himself had discussed in detail in his publications (Zipf 1935). Here, I will rather introduce the original Simon’s model adapted to the framework of city growth. In practice, this just implies a change in the vocabulary employed to express the dynamical rules that define the model.

Simon’s model describes the evolution of a population divided into well-defined groups –the cities. We characterize this division by means of the quantity $m(n,t)$, which gives the number of cities with exactly n inhabitants at time t . This quantity is closely related to the probability distribution $P(n,t)$ introduced in Section 2. In fact, we have $m(n,t) = M(t)P(n,t)$, where $M(t)$ is the total number of cities in the system at the same time. Instead of using the real time t , Simon’s model proceeds by discrete steps, which are identified by means of a discrete variable $s = 0,1,2,\dots$. Each step corresponds to the time interval needed for the total population to increase by exactly one inhabitant. The actual duration of an evolution step –which is determined by a balance between birth and immigration on one side, and death and emigration on the other– is irrelevant to the model. The growth of the population in real time is a separate problem which can be specified and solved independently. As for the model, thus, the elementary evolution event is the addition of a single inhabitant to the total population. Accordingly, the quantity $m(n,t)$ is replaced by $m(n,s)$, the number of cities with exactly n inhabitants at step

s . The starting point of the evolution is given by the initial distribution $m(n,0)$, at $s = 0$, which describes a preexistent population distributed over a certain number of cities.

The evolution is governed by the following stochastic rules, which imply making a decision at each step s , i.e. each time a new inhabitant is added to the total population. (i) With probability α , the new inhabitant finds a new city. In this case, the number of cities M grows by one, and the new city has initially a single inhabitant. (ii) With the complementary probability, $1-\alpha$, the new inhabitant is added to an already populated city. In this case, the destination city is chosen with a probability proportional to its current population. It increases its population by one, and the number of cities M does not vary. Clearly, rule (ii) stands for the multiplicative contribution to the evolution of the individual population of cities. Larger cities have higher probability of incorporating new inhabitants than smaller ones. As it grows, the population is preferentially assigned to those groups which are already relatively large. Rule (i), on the other hand, represents a contribution independent of the preexisting distribution and, thus, stands for additive effects. In particular, it implies that the number of cities grows, on the average, at a constant rate α . Hence, the average number of cities at step s is $M(s) = M(0) + \alpha s$. Meanwhile, since exactly one inhabitant is added per time step, the total population in the system is $N(s) = N(0) + s$.

In order to translate into mathematical terms the evolution rules (i) and (ii), we must take into account the following remarks. First, rule (i) only affects the number of cities with exactly one inhabitant, $m(1,s)$. When it applies, which happens with frequency α per evolution step, $m(1,s)$ grows by one. Second, when rule (ii) applies, which happens with frequency $1-\alpha$ per evolution step, the probability that any city of size n is chosen as destination is proportional to $n/N(s)$ and to the number of cities of that size. Since the chosen city changes its population from n to $n + 1$, this event represents a positive contribution to the number of cities of size $n + 1$ at the next step, $m(n+1, s+1)$, and a negative contribution to $m(n, s+1)$. The contribution to $m(n, s+1)$ will be positive when the chosen destination is a city of size $n - 1$. Summing up these considerations, the average change per step in the number of cities of size n is

$$m(1, s+1) - m(1, s) = \alpha - \frac{1-\alpha}{N(s)} m(1, s) \quad (7)$$

for $n = 1$, and

$$m(1, s+1) - m(1, s) = (1-\alpha) \left[\frac{n-1}{N(s)} m(n-1, s) - \frac{n}{N(s)} m(n, s) \right] \quad (8)$$

for $n = 2, 3, 4, \dots$. These are the equations that govern the evolution of Simon's model in its original formulation (Simon 1955).

From a mathematical viewpoint, Eqs. 7 and 8 are not complicated. First of all, they form a linear system, which can therefore be tackled with a host of well-tested analytical and numerical methods. Moreover, they can be solved recursively. In fact, the solution to Eq. 7 gives the number of cities with one inhabitant.

Once $m(1, s)$ has been found, $m(2,s)$ and, successively, $m(n,s)$ for larger n , are obtained from Eq. 8. The only difficulty is that the equations involve the function $N(s) = N(0) + s$, which depends explicitly on the variable s . Consequently, the system is non-autonomous.

In his original paper, Simon was able to prove that – as I show below – Eqs. 7 and 8 imply a power-law decay for $m(n,s)$ as a function of n . My presentation of the solution will differ from Simon’s in that I will first introduce a continuous approximation to the model equations, replacing the discrete variables n and s by continuous variables η and ξ , respectively. This approximation has the advantage of transforming the infinitely many equations 8 into a single evolution law. The disadvantage is that the new problem is differential, instead of algebraic. Replacing discrete by continuous variables is justified by the fact that, in the distribution of city sizes, we are mainly interested in the range of large values for both n and s , where $m(n,s)$ is expected to vary smoothly. To the first order, we approximate the differences in Eq. 8 by derivatives; for instance,

$$m(n, s + 1) - m(n, s) \approx \frac{\partial m}{\partial \xi}(\eta, \xi).$$

This approximation can be systematically improved by considering higher order terms, as discussed elsewhere (Manrubia and Zanette 2002). The resulting partial-differential equation is

$$\frac{\partial \mu}{\partial \xi} + (1 - \alpha) \frac{\eta}{N(0) + \xi} \frac{\partial \mu}{\partial \eta} = 0 \tag{9}$$

with $\mu(\eta, \xi) = \eta m(\eta, \xi)$. This equation has to be solved for $\eta > 1$, with the initial condition $\mu(\eta, 0) = \mu m(\mu, 0)$ and a boundary condition at $\eta = 1$ derived from Eq. 7, namely,

$$\mu(\eta, \xi) = \frac{\alpha}{1 - \alpha} (N(0) + \xi). \tag{10}$$

I will not discuss the details of the solution method for this linear equation. It is enough to say that, by means of a change of variables (Manrubia and Zanette 2002), the equation reduces to a standard one-dimensional wave equation and is then solved by the so-called “method of characteristics.” In the following, I describe the result in terms of the original variables n and s .

As a function of the population n , the number $m(n,s)$ of cities of size n at step s , solution of the Simon’s equations, shows two distinct regimes. The boundary between both regimes is situated at

$$n_B(s) = \left(1 + \frac{s}{N(0)} \right)^{1-\alpha}. \tag{11}$$

We see that this boundary depends on s , and shifts to higher populations as the evolution proceeds. For populations above the boundary, $n > n_B$, the solution to the continuous approximation of Simon’s model is

$$m(n, s) = \frac{1}{n_B(s)} m\left(\frac{n}{n_B(s)}, 0\right) \quad (12)$$

thus, $m(n, s)$ is directly given by the initial condition $m(n, 0)$. As a matter of fact, this regime can be seen to encompass those cities that were already present at $s = 0$. In Simon's model, preexisting cities – not unlike the oldest cities of real urban systems – are those that reach the largest sizes, i.e. those that are assigned the lowest rank values in Zipf's analysis. We realize that, as advanced in Section 3, information about the initial state of the urban system is stored in the size distribution at the lowest ranks. A detailed study of the effects of the initial condition in the large-size regime, referring to the discrete equations 7 and 8, has been presented elsewhere (Zanette and Manrubia 2001).

In the range of small populations, $n < n_B$, the solution to the continuous approximation of Simon's model is

$$m(n, s) = \frac{\alpha}{1 - \alpha} (N(0) + s) n^{-1-1/(1-\alpha)} \quad (13)$$

This regime encompasses the cities founded during the evolution of the urban system, corresponding to higher rank values in Zipf's analysis. We see that their size distribution follows a power law with exponent $\gamma = 1/(1 - \alpha)$ (cf. Eq. 2). According to Eq. 6, the Zipf exponent is

$$z = 1 - \alpha \quad (14)$$

Since, being a probability, α is positive and lower than one, this formulation of Simon's model predicts a Zipf exponent $0 < z < 1$. The characteristic value $z \approx 1$ is obtained for very small α , i.e. *when the frequency of city foundation is very small as compared with the growth rate of the population*. As advanced in the Introduction, this conclusion is in full disagreement with the bibliographic quotation given there.

In summary, the main results obtained in this section for the original version of Simon's model, within the continuous first-order approximation, are the following. At any evolution stage, the distribution of city sizes shows two well-differentiated regimes. For large cities, which correspond to low rank values, the distribution depends sensibly on the initial condition. This range keeps information on the early state of the urban system and, thus, results to be specific for each realization of the model. On the other hand, the size distribution of small cities, within the range of high rank values, exhibits a universal power-law decay whose exponent is completely determined by the rate α at which new cities are founded. The respective Zipf exponent is always less than one, and the limit $z = 1$ is approached when α is vanishingly small. The two regimes, whose mutual boundary recedes towards high populations as the evolution proceeds, can be immediately identified with two of the three regions of rank plots, described in Section 3. The cut-off region, on the other hand, remains unexplained by this version of Simon's model. Moreover, as presented in this section, the model is not able to produce

Zipf exponents larger than one (cf. Fig. 2). Some of the generalizations discussed in the next section are aimed at alleviating these limitations.

5 Extensions of Simon's Model

It is clear that, in the original formulation of Simon's model, both rules (i) and (ii) involve strong assumptions on the parameters that govern the evolution of the urban system. Specifically, rule (i) establishes that the rate at which new cities are founded is constant, i.e. does not vary with time. Rule (ii), in turn, makes a concrete hypothesis on the size dependence of the probability for a city to be chosen as destination for a new inhabitant. Not without reason, it may be argued that these assumptions are unrealistically simple. Reinforcing this impression, I have just shown that Simon's model is not able to predict some basic features in the rank plots of real urban systems, such as the cut-off at high ranks and the possibility that the Zipf exponent is larger than one.

It has to be understood, however, that the assumptions implicit in the evolution rules have been introduced by Simon, mainly, to facilitate the analytical treatment of the equations and to show, as straightforwardly as possible, that a couple of elementary mechanisms are enough to explain the occurrence of power-law distributions and Zipf's law. If one intends to be more realistic, those strong assumptions can be immediately relaxed, without inherently modifying the basic dynamical processes that define the evolution. In this section, I present a small collection –by no means exhaustive– of generalizations of Simon's model, based on relaxing the evolution rules. Some of these extensions have already been introduced in the literature to solve the above discussed limitations of the model, regarding the detailed prediction of Zipf's rank distributions. My main aim is, nevertheless, to emphasize the flexibility of Simon's model as for possible extensions towards a more realistic description of city growth.

5.1 Time-Dependent Rate of City Foundation

A straightforward generalization of Simon's model consists in assuming that the probability α of foundation of a city when a new inhabitant is added to the system depends on time. Indeed, it is expected that the rate at which new cities appear in a urban system decreases as the total number of cities grow. In the model, a time-dependent rate of city foundation amounts at admitting that α depends on the variable s . In this way, $\alpha(s)$ gives the probability per evolution step that a city is founded by the inhabitant added at step s .

To mathematically implement this generalization, we do not need to rewrite the evolution equations. It is just enough to take into account that, in Eqs. 7 and 8, the parameter α may depend on s . In principle, there are no limitations on the functional form of this dependence. Of course, however, whether the resulting evolution equations are analytically tractable and whether they produce a power-law di-

tribution is a matter to be ascertained in each particular case. In any case, the problem can be dealt with by numerical means.

As an illustration, I consider here a phenomenological model for the time variation of α put forward in the framework of word frequencies in language (Montemurro and Zanette 2002; Zanette and Montemurro 2005). In this model, α decreases with s as a power law of the form

$$\alpha(s) = \alpha_0 s^{\nu-1} \quad (15)$$

where α_0 is a constant, and $0 < \nu < 1$. In the problem of city growth, this form of $\alpha(s)$ implies that the total number of cities increases slower than linearly, $M(s) \approx s^\nu$, instead of displaying linear growth as in the original version of Simon's model. In the relevant limit where $\alpha(s) \ll 1$ for all s – a condition which is insured if the constant α_0 is very small – it is possible to find the solution of the first-order continuous approximation, Eq. 9. As a function of the population n , the resulting distribution $m(n, s)$ shows again two regimes. As in the case of constant α , the large-population regime is determined by the initial condition and, thus, bears information on the initial state of the urban system. The small-population regime, in turn, corresponds again a power-law distribution, but its exponent has changed:

$$m(n, s) = \alpha_0 N(s) \left(1 + \frac{s}{N(0)} \right)^\nu n^{-1-\nu} \quad (16)$$

The associated Zipf exponent is (cf. Eqs. 2 and 6)

$$z = 1/\nu \quad (17)$$

Since $\nu < 1$, we have $z > 1$. We conclude that allowing the rate of city foundation to depend on time, the restriction in the resulting Zipf exponent can be removed. The effect of more general forms of $\alpha(s)$ may be assessed numerically.

5.2 The Cut-Off Regime

Another extension of Simon's model makes it possible to predict the presence of a faster population decay for high ranks, thus providing a plausible explanation for the cut-off observed in the rank plot. Here, I limit myself to a semi-quantitative description of this generalization, as technical details have already been given elsewhere (Montemurro and Zanette 2002; Zanette and Montemurro 2005).

The generalization is based on a realistic consideration regarding the foundation of cities as new inhabitants are added to the population. It can be argued that a single inhabitant is not enough to define the existence of a new city. Rather, there should be a minimal population for a city to enter the regime where the multiplicative process of Simon's rule (ii) acts. This effect can be implemented by modifying the probability that a newly founded city is chosen as destination by new in-

habitants. Namely, the probability that a city of size n is chosen by a new inhabitant can be taken to be proportional to $\max\{n, n_{\min}\}$, where n_{\min} is the threshold population. In this way, a given city must attract n_{\min} new inhabitants before multiplicative growth begins to act. Until then, the probability that the city is chosen as destination is a constant. Note that the threshold n_{\min} may be different for each city. Within this extension, the cut-off of Zipf's plot is interpreted as corresponding to those cities whose size has not yet attained the threshold.

This generalization of Simon's model has originally been introduced in the framework of word frequencies (Montemurro and Zanette 2002). Numerical simulations of the model with an exponential distribution for the value of n_{\min} assigned to each city, combined with an s -dependent probability α of the type discussed in Section 5.1, have provided excellent fittings of Zipf rank plots for several texts in different languages. In view of these encouraging results, it would be interesting to try these combined extensions of Simon's model for city size distributions.

5.3 Size-Dependent Choice of the Destination City

As mentioned above, rule (ii) in the original formulation of Simon's model involves the very special assumption that the probability for a city to be chosen as destination by a new inhabitant is proportional to its size. In other words, the specific growth rate of cities per time step—relative to their current individual populations—is constant all over the system.

This assumption can be relaxed supposing that the probability that a city receives a new inhabitant is not proportional to its population n , but to a function $\phi(n)$. If $\phi(n)$ grows with n faster than linear, the specific growth rate of large cities will be higher than for small cities, and *vice versa*. In the original formulation of the model, one has $\phi(n) = n$. The function $\phi(n)$ stand thus for a nonlinear effect in the multiplicative process. The probability that a city of size n is chosen as destination is given by the ratio $\phi(n)/\Phi(s)$, where the normalization factor is given by

$$\Phi(s) = \sum_{n=1}^{\infty} \phi(n)m(n, s) \quad (18)$$

This normalization insures that the sum of the probabilities over the whole ensemble of cities equals one. In the original model, the normalization factor equals the total population, $\Phi(s) = N(s)$.

Within this generalization, the model evolution equations read

$$m(1, s+1) - m(1, s) = \alpha - (1 - \alpha) \frac{\phi(1)}{\Phi(s)} m(1, s) \quad (19)$$

for $n = 1$, and

$$m(1, s + 1) - m(n, s) = (1 - \alpha) \left[\frac{\phi(n-1)}{\Phi(s)} m(n-1, s) - \frac{\phi(n)}{\Phi(s)} m(n, s) \right] \quad (20)$$

for $n = 2, 3, 4, \dots$. In the first-order continuous approximation introduced in Section 4, they transform into

$$\frac{\partial \psi}{\partial \xi} + (1 - \alpha) \frac{\phi(\eta)}{\Phi(\xi)} \frac{\partial \psi}{\partial \eta} = 0. \quad (21)$$

with $\psi(\eta, \xi) = \phi(\eta)m(\eta, \xi)$. As in the original model, this equation has to be solved for $\eta > 1$, with the initial condition $\psi(\eta, 0) = \phi(\eta)m(\eta, 0)$. The boundary condition at $\eta = 1$ is now

$$\psi(1, \xi) = \frac{\alpha}{1 - \alpha} \phi(1)\Phi(\xi). \quad (22)$$

Now, finally, we have managed to end up with a really complicated mathematical problem. Equations 19 to 21 are very similar to the evolution equations of the original model but, alas, the similitude is only formal. The key difficulty of our new equations for $m(n, s)$ resides in the fact that the function $\Phi(s)$ is generally not known beforehand. In the original model, on the other hand, it coincides with the total population $N(s) = N(0) + s$. Within the present generalization, $\Phi(s)$ can only be given in terms of $m(n, s)$ itself (cf. Eq. 18). Unfortunately, it is not possible to find an independent equation for the evolution of $\Phi(s)$ alone. The distribution of city sizes $m(n, s)$ and $\Phi(s)$ must therefore be found simultaneously and self-consistently.

I have been unable to find a form of $\phi(n)$ allowing me to give an analytical solution either to Eqs. 19 and 20 or to Eq. 21. It seems, not unexpectedly, that the problem must be treated numerically. I leave it open for the reader interested at studying the effects of nonlinear multiplicative processes. To my knowledge, this kind of processes have until now received relatively little attention.

6 Conclusion

The core of this contribution has been devoted to a presentation of the mathematics of Simon's model in terms that, I hope, are accessible to a broad academic readership. To achieve this purpose, I have tried to accompany the mathematical formulation with a qualitative description of hypotheses and assumptions. I have shown that, in its original version, Simon's model is able to explain the occurrence of a power-law regime in the distribution of city sizes, though it fails at predicting some of the Zipf exponents observed in real urban systems, as well as other systematic features resulting from Zipf's rank analysis. The extensions discussed later should have demonstrated that such limitation can be removed—at least, partially—by relaxing some of the assumptions of the model's dynamical rules wi-

thout modifying the key underlying mechanisms. These extensions were mainly aimed at illustrating the potential of Simon's model with respect to possible generalizations in the direction of a better description of empirical data.

Several processes relevant to the evolution of city sizes have not been addressed at all in the presentation of Simon's model. Let me point out three of them. In the first place, I have avoided a detailed description of death and emigration events. I have in fact assumed that the growth of the total population in the urban system is monotonous, the only effect of mortality and emigration being a lengthening of the duration of the evolution step (cf. Section 4). This excludes the possibility that the population might temporarily decrease—a necessary event if one aims at describing, for instance, the eventual disappearance of cities. As discussed elsewhere (Manrubia and Zanette 2002), a separate consideration of mortality and/or emigration implies a change in the Zipf exponent predicted by the model. Secondly, I have not taken into account the possibility of migration flows inside the urban system, between its cities. One can see that a purely multiplicative migration mechanism would exchange population between cities without modifying the city size distribution. On the other hand, additive and nonlinear mechanisms would imply a change in the distribution. This belongs to the class of open problems mentioned at the end of Section 5.3. Third, I have ignored possible events of coalescence of cities which, as indicated in Section 3, shape many modern urban systems. A particularly interesting open problem related to such events regards the persistence of Zipf's law beyond the formation of urban agglomerations. A model for this persistence may shed light on the statistics of the coalescence process itself.

Finally, it is obvious that I have made no attempt to produce a quantitative fitting of real data from city size distributions with Simon's model or any of its extensions. On the other hand, very good fittings have been reported for distributions of word frequencies in language (Montemurro and Zanette 2002; Zanette and Montemurro 2005), musical notes in Western compositions (Zanette 2006), and surname abundance (Manrubia and Zanette 2002; Manrubia et al. 2003), all of which share the dynamical basis of multiplicative processes. It would be nice if this work elicits similar initiatives in the statistical study of urban systems.

Acknowledgments

Collaboration with Susanna Manrubia and Marcelo Montemurro is gratefully acknowledged.

References

Gabaix X (1999) Zipf's Law for Cities: An Exploration. *Quart. J. Econ.* **114**: 3: 739-767

- Gabaix X, Ioannides Y (2004) The Evolution of City Size Distributions. In: Henderson V, Thisse JF (eds) Handbook of Urban and Regional Economics. Vol. 4, North-Holland, Amsterdam, pp 2341-2378
- Gibrat R (1931) Les inégalités économiques. Librairie du Recueil Sirey, Paris
- Manrubia SC, Derrida B, Zanette DH (2003) Genealogy in the area of genomics. *Am. Sci.* **90**: 158-165
- Manrubia SC, Zanette DH (1999) Stochastic multiplicative processes with reset events. *Phys. Rev. E* **59**: 5: 4945-4948
- Manrubia SC, Zanette DH (2002) At a boundary between biological and cultural evolution: The origin of surname distributions. *J. theor. Biol.* **216**: 461-477
- Montemurro MA, Zanette DH (2002) New perspectives on Zipf's law in linguistic: From single textes to large corpora. *Glottometrics* **4**: 86-98
- Portugali J (2000) Self-Organization and the City. Springer, Berlin
- Simon HA (1955) On a class of skew distribution functions. *Biometrika* **42**: 425-440
- Sornette D (1998) Multiplicative processes and power laws. *Phys. Rev. E* **57**: 4: 4811-4813
- Sornette D (2000) Critical Phenomena in Natural Sciences. Chaos, Fractals, Selforganization and Disorder: Concepts and Tools. Springer, Berlin
- Willis J, Yule G (1922) Some statistics of evolution and geographical distribution in plants and animals, and their significance. *Nature* **109**: 177
- Zanette DH (2006) *Musicae Scientiae*. In press, cs.CL/0406015.
- Zanette DH, Manrubia SC (2001) Vertical transmission of culture and the distribution of family names. *Physica A* **295**: 1-8
- Zanette DH, Montemurro MA (2005) Dynamics of text generation with realistic Zipf's distribution. *J. Quant. Linguistics* **12**: 29-40
- Zipf GK (1935) *The Psycho-Biology of Language*. Houghton-Mifflin, Boston
- Zipf GK (1949) *Human Behaviour and the Principle of Least-Effort*. Addison-Wesley, Cambridge

Annex: Poster Sessions

The Features and Challenges of the Urbanization Course in China

Tang Hui Yi, Lu Ming

School of Architecture, Harbin Institute of Technology, Heilongjiang, China

1 The Fundamental Features of China's Urbanization

During the founding of the People's Republic of China in 1949, she is an agricultural country in severe poverty and backward economically and technically. While its urbanization rate is only 10.6% with a total population of about 4.75 hundred million.

China implemented compulsorily a fundamental industrialization under the concentrated planning economic system, which implemented all-round public ownership. During this process, an urban/rural areas separation policy was induced, i.e., rigorously differentiating the status of residents of urban/rural areas through the urban/rural permanent residence registered system. While in the cities an all-round welfare policy was implemented. At the same time, the urbanization development was limited through the national policy. It was advocated to stress on developing small cities, and many industrial enterprises were located in a scattered pattern. In such context, it is formed that the urbanization level was considerably lagged behind the industrialization level; and the urban infrastructure construction was in shortage. Through the 10 years turmoil of the "cultural revolution", the urban planning and construction had been left in a state of stagnant and discard.

Since 1978, the nation began to try to implement the policy of reformation and open door. At the starting point of 1980, the country's average urbanization rate jumped to 20.6%.

Since then, a rigorous family plan had been brought into practice, this had resulted a reduction of population explosion by about 3 hundred million. And a nation wide systematic urban and regional planning had been implemented for all cities and regions. The urbanization process had been brought under control in some extent. The economic reformation starts from the rural area. The system of "people's commune" was abolished; implement instead a system of "contracted responsibility linking remuneration to output". After the liberation of the initiation of the farmers, they began to engage in various non-agricultural industries, including industry, commerce, transportation, building and service, etc. They took advantage of various conditions to create town firms, and promoted the vigorous development of towns.

The government supported the development of town firms and towns, because it solved considerably the problem of employment in which the rural exceeding manpower transfer to the non-agricultural industries; and at the same time, it complemented the problem of lag of China's urbanization process. However, the distribution of the town firms is too dispersed and hence induced the severe environmental pollution of the rural area. The infrastructure and social service facilities was in an extreme shortage, the civilization level of the people there was lower, so the quality of many towns was on a low level. This relies on the readjustment of its layout and the assistance of the central cities.

In such development, the Chinese urbanization system appeared in two subsystems, i.e., the principal urbanization (pure urbanization) subsystem and the rural town movement (quasi urbanization) subsystem. The former was developed with domestic and foreign investment, that's the conventional urbanization; it is a top-down process; while the latter is the developing towns mainly formed with the rural pushing forces, that's a bottom-up process.

From the point of view of planning and orderliness, the former were planned more comprehensively while the latter lack of planning intervention and mainly formed spontaneously. Chinese urbanization developed rapidly with a rapid economic growth of average annual GDP increasing about 9% during the last two and one half decades. Up to the end of 1999, China had 668 cities with big, media or small scale and 19'875 towns (including county centers, central towns and organic towns), while the national average urbanization level is 30.6%.

But the number of towns appears a tendency of decreasing in recent years because of the concentration process; for example, in 2002 the total number of towns had reduced by 15% in comparison with those of 1997.

Since 1978 through 2003, the number of ultra-big city with population greater than 1 million increased from 13 to 49, the number of big city with population between 0.5 and 1 million increased from 27 to 78, and the number of medium size city with population between 0.2 and 0.5 million increased from 59 to 213. While several megalopolises emerged in China with the larger ones as Yantze River Delta, Zhujiang River Delta and Around Bohai Sea Area. The up-grad of people's consumptive structure gave impetus to the modernized up-grad of industrial structure.

Up to 2005, the national urbanization level reached 43%, with national average GDP per capita exceeding 1'700 U.S. dollars. Generally speaking, the current process of Chinese urbanization has already begun to reach a stage of rapid development. However, the urbanization level is still lagging behind the industrialization level. The rural population still constitutes the major part of the total population. If without counting in those who came from the rural area and seeking for jobs in the cities (its number is about 1.2 hundred million), then the rural population would be 6.5 hundred million, about one half of the total population. Now a day, the way of operation of agriculture in China is still mainly the household management with mechanization level not high enough (with agricultural machine mobile team operating during the spring ploughing season and harvesting season in the fall in the major grain production areas), while the large-scale managing ag-

riculture only constitutes a small part. If in the future it develops to the intensive farming as a major way of operation, then the exceeding manpower of the rural area would be released even further by 5~6 hundred million! At present the tendency of Chinese urbanization is still concentration. The development of towns has an attribute of being forced under the “binary social structure” (which had begun to ease during recent years): It was pushed out from the rural area with lower civilization development rather than dispersing from the central cities as higher developed civilization because of crowding. Now the higher developed economic or cultural factors still tend to concentrate toward the central cities.

The industrial transition toward higher level at present is still occurring in central big cities. However, the tertiary industry as well as higher level ones now is in a state of underdevelopment from meeting the needs. So that it is now the big-city-civilization as the representative despite a lot of small towns is developing and being encouraged.

As for the administration level of the urban development, under the situation of the transitional period from the central planning economy toward the market economy, the urban government is changing its way of operation in a reformation from the state of over intervening the economy and decision-making toward the orientation of modern democratic and service administration.

2 The Encountered Challenges of the Virtuous Development of China’s Urban and Rural Systems

- The land of China is in a considerable shortage despite she has a territory area of 9.6 million square kilometers, roughly the same as those of the U.S. However, its plain area is only 12% of the total in comparison with 70% in the U.S. while the population is 5 times over it. The average arable land per capita is 0.11 acre, less than one half of the world average value, and it is decreasing annually because of the occupation of construction and the developing of deserts; on the other hand, the population is increasing by over 10 millions annually. For this reason, the nation has implemented the most rigorous policy and measures to control the arable land.
- The un-renewable resources and energy (including petroleum, gas, coal, iron, copper and aluminium, etc.) is exhausting rapidly because the rapid development of economy. To seek for alternate energy and resources is an urging task.
- The shortage of fresh water resource in China has been realized in clarity, its average occupation quantity is 2 400 tons per capita, less than one half of the world’s average value; further more, the precipitation distribution is not in equilibrium: 81% concentrated in south of Yangtse River area, only 19% is in north of it. So that about 400 cities mostly at north of Yangtse River is in a shortage of water supply. This had an impact on the production and people’s living of the relevant cities. The nation had started the huge engineering of “transporting the southern water to the north”. However, it is predicted to accomplish in 2030.

- The trend of deterioration of China's eco- system and environment had not yet reached a fundamentally turning point despite of great efforts had been made by the government and large amount of investment had been put in. large scale of vegetation, afforestation, improving soil, and pollution control is undertaking. It is planned to get to the turning point in 2010.
- The problem of employment of urban and rural population is one of major tasks of the government in the coming decades, including the "post-off workers" in the cities because of the reformation process of the state-owned enterprises, and the exceeding manpower of the rural area which had been counted as about 2.7 hundred million last year, among them, about 1.2 hundred million is getting into the cities to find jobs. However, in the cities, the infrastructure, wage payment, education, medicine and legal systems, etc., —the relevant service system for them is vulnerable and urging to improve. The rich-poor gap is broadening today in China, especially the income gap between urban and rural areas broadened rather quicker than the else. The income distribution mechanism is expected to be improved.
- The problem of agriculture, villages and farmers as a whole is one of protruding problems encountered by the urban-rural development of current China. In order to promote China's urban-rural development including the "dual track urbanization system" onto a virtuous track, it is expected, so, to propel a top-down consistency movement: i.e., fully play the radiation role of the central cities, to propel the 6th industrial revolution (i.e., the revolution of the grand agriculture), push the factors of modern science and technology, economy, culture and education to the small towns, with which to promote the modernization of the grand agriculture (including agriculture, forestry, deserticulture, grassiculture and sea-management), and the poverty-off and getting rich of the broad masses of farmers, herdsmen and foresters and at the same time, to promote the improvement of the eco-system and environment.

Introducing to Simulating Methodology of Urban Systems

Tang Hui Yi, Lu Ming

School of Architecture, Harbin Institute of Technology, Heilongjiang, China

1 Open Complex Giant Systems (OCGS)

With many categories of subsystems and with multi-layered structure and with complex relationships in between them, furthermore, it is opening to the environment.

An even higher category is the *open special complex giant systems* (OSCGS), as for example, the social system. With human being as the subject of its subsystems which might include also many sorts of intelligent machines.

1. The system and its subsystems have respectively various sorts of information communications with the environment
2. Each subsystem acquires knowledge through study.

Due to the role of human willingness, the relationship between subsystems are not only complex but also very easy to change with time according to situation. This is the open special complex giant system which not only constitutes of a lot kinds of subsystems but also with knowledge playing an extremely important role.

Its complexity may be described as the following:

1. communicate between subsystems with a variety of ways
2. with many kinds of subsystems each with its qualitative model
3. with different knowledge expressions and different ways to acquire knowledge in each subsystem
4. the structure of subsystems might change with the evolution of the system, so the structure of the system will change successively.

The study of man as a complex giant system may be seen as the microscopic-study of the social system. On the aspect of macroscopic-study of social system, according to Marxist concept of social formation, there are 3 *social formations* in a society, e.t.,

- Economical social formation
- Political social formation
- Ideological social formation

The social system can be divided into 3 *constitutional parts*:

- Social economic system
- Social political system
- Social ideological system

And there are 3 relevant *civilization constructions*:

- Physical civilization construction
- Political civilization construction
- Civilization construction

Its practical instrument is *social system engineering*. The social system engineering is the organizing and managing technique to induce the harmonic development among these 3 subsystems and between the social system and its environmentits 3 subsystems are:

- Economical system engineering
- Political system engineering
- Ideological system engineering

The OCGS involves biology, thinking science, medicine, earthology, astronomy and social science, etc., so it is a very broad research area. Its study will not only promote the theoretical development of these varieties of disciplines, but will also explore a new perspective for the mutual break-through in between these theories.

2 The Methodology of Research for the OCGS for OCGS

There is still yet no theory analyzing from microscopic to macroscopic point of view, no statistic mechanics theory constructed starting from the interactions between subsystems. Somebody tried to use those methods for dealing with simple giant systems to deal with OCGS. They didn't see the limitation and the application domain of these methods; as in game theory, they over simplified human sociality, complexity and the uncertainty of human psychology and behavior. Similarly, to use system dynamics, theory of self-organization, etc., to the study of OCGS, one can't get success, the cause is also the same.

Others raised the problem of OCGS to the height of philosophy.

Somebody framed out the so called "theory of universe hologram unification".

Somebody even advocated that "the part contains the hologram of the whole", "the part is the whole, and vice versa, the two are absolutely the one".

The practice had proved that, in dealing with OCGS, (including social systems) effectively, now the unique usable method is the *metasynthetic method with qualitative and quantitative processing in combination*.

In the research and application practice, usually it is necessary to combine scientific theory, experience and experts' judgments together, work out empirical hypothesis (judgment or guess); this empiricall hypothesis cannot be proved rigorously with scientific ways, usually it is qualitative cognition, but its reality

could be checked with empirical data and model with hundreds or thousands parameters; and these models should be constructed on the basis of experience and realistic understanding of the system, through quantitative calculation and comparisons time and again, finally achieve the conclusion. Such conclusion is the best one that we can reach at current stage in recognizing the objective reality, that is the cognition risen from qualitative to quantitative ones.

Generally speaking, the metasynthetic method with qualitative and quantitative processing in combination is, in its essence, to combine organically the experts group (with relevant different disciplines), data and various information and the computer technique, to combine the scientific theories in different disciplines with human empirical knowledge. These three aspects in turn constructed as a system. Taking advantage of the integrated and synthetic potentials of this system, that's why this method can be applied successfully.

In recent years, some foreign scholars suggested the "meta-analysis" method, dealing with information of various areas to make cross-area analysis and synthesis, but it has not yet got matured and with relatively simpler method.

3 Metasynthesis May also Use Knowledge Engineering

In the process of solving problems with the method of *metasynthesis*, the empirical knowledge of experts groups and experts played an important role. If we goes further, to insert knowledge engineering instead in some extent of men's role, then it relates to the expression and treatment of knowledge.

Knowledge engineering is an important branch of artificial intelligence. The expert system is a sort of typical knowledge-type system.

A system constitutes of men, expert system and intelligent machine, as subsystems must be a men-machine interact system.

All subsystems mutual coordinate harmoniously with men conducting and making decisions at the key points and with machine doing the replicating and heavy works. Men and machine communicate with various convenient ways.

The methodology of the research of knowledge-type systems had been probed in recent years.

The key problem in knowledge engineering is the expression of knowledge, that's to express various kinds of knowledge into a form that can be accepted and treated by the computer.

The feature of the knowledge-type system is to solve problems with a heuristic way controlling by knowledge, not precise quantitative treatment, not quantitative mathematical model, but qualitative method, or qualitative combined with quantitative ones to do the system synthesis.

Qualitative modeling is a method coding with deepening knowledge, concerning only the tendencies of variations as increasing, decreasing, or constant, etc.

The *qualitative reasoning* is to operate on the qualitative model, in order that the behavior of the system is got or predicted. The stressing point is the description of its structure, behavior and function.

There were three representative works in this area done in the U.S., e.t.,

- the component centered model by De Kleer, a qualitative physical system that can be used for explaining and prediction.
- the constraint centered model by Kuiper.
- the process centered model by Forbus.

The motive in the research of qualitative modeling and reasoning in the *knowledge engineering* is the research of conventional knowledge, its expression, storage and reasoning. In reality, many important works in *artificial intelligence* was considered from the point of view of systems. Someone advocated that the research of artificial intelligence can be summarized as the discipline to research the calculation method of acquisition, expression and usage of various qualitative Models (physical, conceptual, cognitional or social system model).

Now the thought of metasynthesis in the artificial intelligence area had been drawn attention. Generally speaking, after we broadened the concept of "opening" and "complexity" of systems, we could have a deeper recognition to systems and a broader including contents. After we explained that the OSCGS belongs to the highest level in the classification of systems, we have in reality made a break through between the areas of system science and artificial intelligence. So that the various kinds of intelligent systems characterized with knowledge then all belongs to an united, clarified scope. This will benefit for establishing the theoretical basis of OCGS. This is the unavoidable result of the development of current science.

4 The Meaning of the Research of OCGS

We can see that, the method of metasynthesis with qualitative and quantitative processing in combination has in summarization the following features:

1. Based on the features of the OCGS that has complex mechanism and with a plenty of variables, so one combines organically the qualitative study with the quantitative study, rising from the qualitative cognition of multi-aspects to the quantitative cognition.
2. Because of the complexity of systems, it is necessary to combine scientific theories with empirical knowledge, to synthesize and concentrate the scattered knowledge of men of the objective things, in order to solve problems.
3. According to the thought of systems, to combine multi-disciplinary theories in research.
4. According to the layered structure of OCGS, unite its macroscopic study with microscopic study.

Exactly these features make it possible that this method has the ability to solve complicated problems in the OCGS. For this reason, it has a very important meaning.

The object that modern science and technology is probing and studying is the *objective world as a whole*, however, modern science and technology had been divided into different disciplines. The OCGS had been generally appearing in this world as well as in the universe, so its study has a general meaning.

However, the scientific theories in the past cannot solve the problems of OCGS, there are causes which can be looked for in the history. As it is well known, scientists from different areas found that in life systems and non-life systems seems appear entirely different laws, during the 1960s, J. Prigogine and H. Haken in their research defined a new time direction: *from chaos toward order*.

The achievements of modern science during recent decades are very important. However, the success of the theory of dissipative structure and synergetics also made many people over optimized. Although scientists had made great efforts to realize the transition of social science from "descriptive science" to "precise science", and had acquired achievements, the social science system as a whole still has yet a long way to be a "precise science".

From the above mentioned discussions we can see that the OCGS and its research method, in reality, is to concentrate the large amount of scattered qualitative cognition, small drops of knowledge, even the mass'opinions into an integrated structure, to reach a quantitative cognition, is from incomplete qualitative to relatively complete quantitative, that's a leap from qualitative to quantitative.

Of course, through such study in one aspect, got a plenty of accumulations one can rise again to the qualitative cognition of the system as a whole, and reach a higher level of cognition, form another cognitional leap.

5 Complementary Remarks

For dealing with the OCGS, Academician Tsien Hsue Shen advocated the method of metasyntesis, or metasyntetic engineering, its kernel part is "modern science and technology system" and the "hall for workshop of metasyntetic engineering". To develop it theoretically as a school, it is "metasyntesis"-during the IT era, a school for men to acquire wisdom and intelligence, to enhance their capability to recognize and reform the world. This is also an important contribution to the development and deepening of Marxist philosophy. "Metasyntetic wisdom" take advantage of modern information web, with men-computer in combination in which men play the main role, concentrate the relevant experience, knowledge and wisdom of all time and places. Its feature is the web wisdom formed by sinking in the broad cyber-space, it is a new type of way of thinking and thought system during the era of knowledge blast and information tide. The kernel of "metasyntetic wisdom" is the combination of science with philosophy.

The human intelligence can be divided into two parts: the characteristic intelligence and the quantitative intelligence. The former can be nurtured through arts, music or painting; while the latter can be nurtured through analysis, mathematics or physics training, and it can be realized with the help of computers, that is the part of human logical thinking.

The "characteristic intelligence" is a sort of wisdom that handling the global aspect technically from a qualitative, macroscopic point of view. The "quantitative intelligence" is a sort of wisdom for quantitative, microscopic analysis, summing up and reasoning.

The "man-computer combination" with man plays the main role, where the "computer" does not alternate the "man", but assists him.

Considering from the point of view of information treatment, to make the people's "characteristic intelligence" and "quantitative intelligence" combining with the "high capability" of information treatment of the computer, achieve the mutual complement of the qualitative (non-precise) and quantitative (precise) treatments.

In the process of solving complicated problems, the works which can be formalized is to be put to the computer as much as possible, while the key point and those which can't be formalized is to rely on people's direct participation or indirect role.

In this way, not only the extremely complex problems can be treated, but also can concentrate wisdom to an utmost extent.

Firstly give some qualitative solving thinking and ways (qualitative judgment), and then do the quantitative analysis with the computers. Tsien Hsue Shen further proposed to establish a "hall for workshop of metasynthetic engineering". This is a huge intelligent system with man-computer in combination.

If it establishes on the basis of internet, not limited in time and space, then the "hall for workshop" becomes a "cyberspace". This would be the embodiment of dialectic thinking and a "working hall" of the 21 century.