Models in Spatial Analysis

Models in Spatial Analysis

Edited by
Lena Sanders

iSTE

First published in France in 2001 by Hermes Science/Lavoisier entitled "Modèles en analyse spatiale"
First Published in Great Britain and the United States in 2007 by ISTE Ltd

ISTE Ltd
6 Fitzroy Square
London W1T 5DX
UK

ISTE USA
4308 Patrice Road
Newport Beach, CA 92663
USA

www.iste.co.uk

Library of Congress Cataloging-in-Publication Data

Modèles en analyse spatiale. English
Models in spatial analysis/edited by Lena Sanders.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-905209-09-5
1. Geography--Mathematical models. 2. Spatial analysis (Statistics)--Mathematical models.
3. Geomatics. I. Sanders, Lena. II. Title.
G70.23.M513 2007
910.01'5195--dc22
2007012050

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 13: 978-1-905209-09-5

Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire.

Table of Contents

Preface

This book offers a relatively large overview of modeling *families* which exist in spatial analysis and of the different mindsets that went into them. Models presented are related to space organization principles, localization logic, the form of spatial interactions[1] and their role in space dynamics. The expertise of the authors is thematically (geography, economy, hydrology, ecology, etc.) and methodologically (mathematics, statistics, computer science, physics) varied, and several authors use an interdisciplinary approach.

You will notice two types of chapters, some proposing a view on spatial analysis from a thematic point of view, the others focusing on a key concept or on a specific methodological field of spatial analysis. In both cases, the goal was to propose a relatively large overview of the question and also to support the presentations with the help of concrete application examples. A system of cross-referenced reading was adopted for the French edition of the book, which has enabled exchanges between the authors and has facilitated correlation between the chapters.

We wish to thank the group of authors who have graciously accepted to collaborate on this book, the publisher for the translation work, François Durand-Dastès for his help in revising proofs, as well as Eugénie Dumas and Malika Madelin for their precious help in completing this project.

Lena SANDERS

---

[1] An approach illustrated by the cover image, a visualization tool developed by A. Dyèvre and A. Subiron to represent networks of cities in the SimPop model.

Introduction

Models in Spatial Analysis

Which model and which spatial analysis?

The spatial dimension plays a key role in many social phenomena. On the one hand, things are unequally distributed through space, creating spatial differentiation, segregation and discontinuities. On the other hand, there is a feedback loop between a society's organization and a space's configuration. Most of the time, the modeling of spatial phenomena and processes requires the combination of knowledge and skills from various fields, some involving the theme of research, others originating from computer science, statistics, physics or mathematics. These latter subjects are relevant because they provide stimulating methodological prospects to theorists interested in the organization of space and the evolution of its structures (whether they be geographers, urban planners, archaeologists, ecologists, agriculture scientists, etc.). Besides, they also offer useful technical frameworks for formalizing the thematician's theoretical models. On the other hand, the latter provide problems and data that can be used to implement and test models designed by mathematicians, statisticians and computer scientists. In such a context, the phrase *spatial model* takes on different meanings. Even though researchers in different fields may agree on a relatively broad definition of the concept of a "model" as "a schematic representation of reality, developed with the goal of understanding and explaining it"[1], the referents are significantly different from one subject to another.

Introduction written by Lena SANDERS.

[1] A definition close to that of P. Haggett [HAG 65] used by François Durand-Dastès in Chapter 1.

In order to illustrate the diversity of meanings that come with the word "model" in a given application, in the following sections we will rely for the most part on two concrete examples. In the first, the goal is to model the evolution of the spatial distribution of the population over a few decades and, in the second, of the land use . From a methodological perspective, each of these two examples involves dealing with and modeling "spatial dynamics". However, *modeling* spatial dynamics can take on several different meanings: it can consist of describing changes as clearly as possible, or of finding causality underlying the type and speed of the observed evolution. These two approaches, the former used to describe and the latter to explain, should logically complement each other and be conducted one after the other. In practice, however, the line between the two is not always clear, and the structure of the information itself leads to different attitudes in research. If this structure is complex, the first step can constitute the entire work of modeling. Indeed, the objective is then to provide a simplified picture, in the form of a map, for example, of a complex information obtained from several sources. If, on the other hand, the information is described by a simple and relatively regular structure, such as in the case of statistical information, the first step is short and most of the modeling work is done in the second. In order to study spatial differentials in population growth, the geographer will build an explanatory model based on mechanisms he knows are fundamental, the role of rent, the effects of competition for space, urban polarization, the deterring effect of distance, etc. In order to study the evolution of land use, the agriculture scientist will develop a model involving the modes and structures of farming, the specializations and quality of soil.

In any case, the first step consists of identifying the elementary objects that give the level at which the information has to be collected. These objects can be cells or pixels (image analysis), households or firms (surveys or census results), or spatial entities (morphological or administrative units). When studying a settlement system, the elementary entities can be municipalities or agglomerations. When analyzing land use, plots or farms can be used. This identification becomes more complicated in a dynamic framework, since there can be changes in the object's semantics (when a given agglomeration "absorbs" another, for example), identifier (a plot can be divided in two) or its spatial extent (expansion of a farm) [CHE 95]. Once the objects are identified, we have to come up with ways of measuring changes that make sense in regards to the topic, and which are compatible with the spatial scale. Already, throughout this entire process of formalizing information, we are dealing with *models*. These are, on the one hand, conceptual models concerning the meaning of the chosen indicators, particularly the link between the phenomenon being studied (which refers to a *theoretical* framework) and the set of data obtained from its measurements (which refers to what is *observable*), given the observation levels that were chosen. These are, on the other hand, models of data that are necessary to organize thematic and geometric information in order to process them (relational , object-oriented, vector/raster, etc.).

Once the thematic model has been designed and the gathered information has been organized, it is necessary to formalize the model in order to validate it, to test hypotheses and, sometimes, to establish scenarios to conduct prospective studies or to help decision making. Different approaches are then possible, each one associated with a specific set of skills.

If we choose a geomatics approach, we would develop and use the functionalities of a geographic information system (GIS) in order to connect layers of information related to different aspects of the phenomenon being studied and to find evidence of possible spatial regularities. We can construct, for example, requests designed to pinpoint plots located at a certain distance from a freeway interchange and that have undergone a specific change in land use. If we choose a statistical framework, a typical method consists of identifying the variables that have to "be explained" (dependent variables), quantitative (population growth rate, variations in the number of jobs) or qualitative (changes in land use, changes in the economic profile), and "explanatory" variables that refer to the features of the agglomerations or land plots and express the hypotheses set out in the thematic model. For the statistician the discussion will focus on the choice of a *statistical model*, linear, loglinear, exponential, etc., depending on the statistical nature of the variables and on their relations. Facing the same question, researchers trained in mathematics or physics will develop a dynamic model in the framework of complex systems (for example, a non-linear prey-predator model) that includes bifurcation properties. These will be particularly useful if we consider that the parameters that characterize the general environment of the system being studied are governed by threshold effects that may cause a qualitative change in the behavior of settlement dynamics. Formalizing the model within the framework of computer sciences will lead to the use of cellular automata or multi-agent systems and to focusing on interaction and emergence properties. Such models work by conducting simulations, and these are particularly useful when trying to establish the effect that the set of local interactions has on the evolution of spatial structures at a higher level [WHI 97, BEN 04, POR 04].

To the thematician, these different approaches represent a series of methods that are useful to formalize and test his thematic model, each one leading to different and complementary possibilities. His choice will depend on his abilities, his collaborations or, better, on the adequacy between his hypotheses on the why and the how of the change he is interested in and the properties of the available statistical, physical, geomatics or computer model. In some cases, he may have to use a series of models: a GIS to create spatial indicators; a statistical model to test the relevance of the relationship between some of the variables used, or also to estimate missing information; a differential equation model to deal with the overall evolutions of a certain number of state variables; a cellular automata model to simulate the local spatial interactions; a multi-agent system to account for the effects of a cooperation between several agents. In order to implement a thematic model, it thus becomes necessary to string together, to couple and to integrate models originating from different fields, and confusions of course can occur when the same word “model” is used without being associated with a complement that specifies its meaning.

Each of these groups of methods has its specific advantages and integrates in its own way the spatial dimension of the phenomena being studied. The statistical approach emphasizes the covariation mechanisms of the phenomena, and space can be included in the analysis by means, first, of the type of statistical individuals considered (spatial entities, location pairs) and, second, of the variables (distances, shapes of spacings or types of neighborhoods, for example). In dynamic models, space can, in the same way, intervene as a frame for the studied objects (spatial entities), and also through an interaction term, connecting the dynamics of the entities with each other (the distance or the commuting cost between two entities, for example). If the cellular automata are used to represent geographic space, the very driving force of the change is spatial, since any change can be determined based on the neighborhood configuration. Each of these methodological frameworks can be used with the goal of describing, exploring or explaining. The lines that separate these approaches are not always clear and depend more on how the researcher designs his methodological device than on the methods themselves.

The question can be stated in similar terms for *spatial analysis*. Depending on the contexts and the authors, different meanings are bound to this concept in the case of geography alone. Some give the phrase a strong thematic meaning: “Formalized analysis of the configuration and properties of the geographic space, as it is produced and experienced by human societies” [PUM 97]. Others emphasize the spatial aspect of the phenomenon being studied: “Searching, within the features of spatial entities, what is relevant to their geographic positions, particularly their *relative* geographic positions, making it necessary to model the spatial structure” [CHA 95]. Others, finally, insist on the methodological aspects, and spatial analysis refers to the “set of mathematical and statistical methods designed to specify the nature, the quality, the quantity related to places and to their relations – all of this defining space – by studying both attributes and localizations” [BRU 92], or also, in a similar approach, a “whole cluster of techniques and models which apply formal, usually quantitative, structures to systems in which the prime variables of interest vary significantly across space” [LON 96]. GIS users, on the other hand, when they use the phrase spatial analysis, refer to a set of functionalities, such as superposition techniques, topological and geometric functions, interpolation techniques, network analysis, etc.

Again, there can be some confusion between thematicians, such as geographers, who see spatial analysis as a conceptual framework in which the effects of position, neighborhood and co-occurrence play a key role in helping to make clear the interactions between space and society, associated with elementary spatial models referring to gravity, potential, diffusion and to the tools of spatial statistics (for example, spatial autocorrelation, variogram), and GIS users for whom the phrase spatial analysis sometimes means the same thing as the geometric and superposition functionalities of GISs.

Thus, a computer scientist and a thematician do not share the same meaning for the phrase “models in spatial analysis”. We will not discuss in detail in this book the differences in points of view, but instead we will adopt the following statement. The idea is to describe the approaches concerned with *thematic* modeling, and to give an idea of the diversity of the methodological choices made in practice by those whose object of study is strongly rooted in space. Any modeling requires geographic information to be formalized, but the emphasis does not rest explicitly on this question, but rather on how the hypotheses are formalized. The goal is to understand, explain and simulate spatial organizations, special localizations, the emergence or the end of certain phenomena in certain locations, and the types of interactions between different places.

Geographic information plays a central role in all of the studies detailed in this book, even if there are few explicit references to GISs. Despite going unnoticed in these modeling studies, they have nonetheless a strong implicit presence and are essential tools in some processing steps. Certain spatial analysis methods useful for developing a thematic spatial model exist in the GIS, or can easily be internalized, while others are used in completely independent ways or are coupled. Geographers have produced remarkable advances over the last decade in coupling and integrating tools of spatial analysis, simulation models and GIS ([MAG 06]).

Each chapter in this book shows the relationship between a conceptual framework of spatial analysis and a set of methods and models. The frameworks are related to the questions of localizations and interactions in space, of geographic scales and levels of observation, of spatial dynamics. Some authors favor the contribution of a group of methods originating from statistics, mathematics, and computer science, while others base their presentations on theories deeply rooted in geography or economics. All of them attach a key importance to the connection between the thematic meaning of their studies and the properties of the tools they have chosen. Space is present both as an object – the objective being to describe, understand and predict localizations, spatial configurations and their evolution – and as a part of the explanation – the objective being to assess the effects of location and neighborhood on these localizations, configurations and changes.

Two central and recurring concepts: spatial interactions, levels of observation

Every chapter in this book deals with situations where space plays a key role, sometimes as a passive frame but most often as an active function. On the other hand, the various chapters are very diverse in terms of the phenomena being studied (for example, water run-off, localization of public services, population mobility), of the analysis tools (for example, multivariate analysis, differential equations, multi-agent systems), of spatial scales (from a square foot cell to the national level) and time scales (from a minute to a century). This book thus illustrates the diversity of questions and practices in spatial analysis modeling. Beyond this diversity, there are recurring concepts that are unavoidable in any approach in terms of spatial analysis. These concepts are spatial interactions and levels.

Modeling spatial interactions

The concept of spatial interactions is used in most chapters, and is in fact at the core of the spatial analysis approach. When the locations have been characterized, when the intensity and the types of spatial differentiations have been determined, and when the similarities and contrasts have been brought to light, the next task consists of finding the relationship between these features of spatial organization and the exchanges these locations maintain among each other, as well as the mutual influences they have on each other. These influences and exchanges originate from the disparities of space; in fact Alain Franc uses the phrase “interactions driven by spatial organization” (Chapter 9). The other way round, these interactions, in turn, contribute to producing disparities. Thus, spatial interactions reflect relations of complementarity and/or competition between the locations and act as a driving force in the transformation and the dynamics of spatial systems. The interaction between these locations can either be the object of research itself (when trying to explain why some flows are more or less significant between certain pairs of locations), or represent a constraint (movement from one location to another is viewed as a constraint on achieving a task for a given agent), or also constitute the explanatory element of a dynamic (the start of or an increase in a phenomenon in a location brought about by exchanges with other locations).

As a result, spatial interactions take on various forms, operate on several levels, and their analyses lead to different formalizations. In the following sections, we will attempt to express the major aspects of this diversity that are dealt with throughout the different chapters.

Exchanges between pairs of locations

In a first set of models, locations have symmetrical roles, because each location has the possibility of being a transmitter and a receiver. When studying interregional migratory exchanges, for example, each considered area represents a possible settling location for migrants from elsewhere and, at the same time, is the departure location for a certain number of inhabitants who have decided to migrate. The most common spatial interaction models are designed to express the intensity of the flows or exchanges between an original location and a destination, depending on the characteristics of these locations (in terms of departure potential and attractiveness of the destination) and on the deterrent effect of the distance between the two locations. François Durand-Dastès (Chapter 1) uses the gravity model, one of the most common in spatial analysis, to account for what he refers to as the "universals" of modeling. In more general terms, the author stresses the connection between interaction and causality, and underlines the importance of considering spatial interaction for the choice of a type of systemic formalization in spatial analysis.

Exchanges between locations can be material or immaterial, they can be expressed in terms of their quantity (number of migrants) or their nature (presence of a connection by air), and they can be described using sagittal, chorematic or matrix models. The combination of all the spatial interactions operating on an elementary level sometimes leads to the emergence of complex spatial structures. Pierre Frankhauser and Denise Pumain analyze the contribution of fractals to the study of such structures (Chapter 10).

Exchanges between locations with different statuses

A certain number of interactions, on the other hand, are dissymmetric from a semantic perspective, and explicitly involve differences in status between the locations that are considered. Such is the case for diffusion phenomena, in which one or several places are considered as favored transmitters (of information, an innovation, a disease) whereas the other places constitute locations for the receivers, the adopters, the contaminated. This is also the case when considering consumers, characterized by where they live, searching for a localized service: school, day care center, hospital, etc. Depending on the problem, different formalizations are possible. The first thing to do is to determine on what geographic level the starting place (the location with access to the innovation or accommodating the population expressing a potential demand) should be considered: building, block, city or a larger area (in practice, an administrative district most of the time). Destination locations can be identified as entities of the same nature, or simply defined by localized points in the case of a service for example. We then have to determine an adequate model to describe the associated interactions and their consequences on the organization of space and on its dynamics.

In Chapter 3, Dominique Peters and Isabelle Thomas present models that can be used to determine the optimal location of public services given the distribution of the population concerned by these services. Rather than using the word interaction, they suggest the concept of *spatial adequacy* between the supply and the demand for these services. Chapter 5, by Thérèse Saint-Julien, focuses on the spatial diffusion phenomena and on how they are formalized, the idea being that a certain number of transformations of space are due to *contagion* effects. The phenomenon may be physical, such as the propagation of a plant or a disease, or social, by means of imitation behavior, in areas as diverse as innovation in agriculture or architecture, and behaviors in terms of fertility. The most common formalizations, developed in the 1950s, remain valid today on a conceptual level, with much broader possibilities for technical applications.

In Chapter 4, Sonia Chardonnel shows how the principles of *time-geography* are used in order to simulate the impact of a change in public transportation on the individual possibilities of achieving space-time trajectories. In this case, spatial interaction is in the form of commutes, for example "home – day care center – workplace – day care center – home", in accordance with certain constraints on the individuals' schedules and how long they have to stay in certain locations. The point of view is that of the individual subject to constraints in space (reaching a destination while stopping at certain required stops, for example, the day care center) and in time (observing the opening hours of the day care center and work hours).

Chapters 3 and 5 illustrate how, faced with the similar problem of access to a service, the model depends on the research prospects: in one case, the objective is to optimize the localization of a set of recycling centers, while in the other, the access to day care acts as a constraint on the home-to-work commutes of individuals. In the first case, an optimization model is used to provide the distribution of recycling centers that leads to minimize the sum of individual trips. In the second, the locations of the home and day care centers are set, and simulations based on logical rules, involving schedules and lengths of time, are used in order to test the impact of improvements on public transportation, in terms of frequency, on the range of individual possibilities.

Influence of a location on neighboring locations

Spatial interactions can also be expressed as an *influence* of a location on another, without being explicitly embodied in the form of a measurable exchange or flow. Cellular automata are often used to formalize the effect of such influences on local change and simulate the spatial configurations that arise at a global level. In Chapter 8, Françoise Dureau and Christiane Weber present a model for land use change at a pixel level. The land use is identified based on the analysis of a satellite picture, then the potential for change in land use is determined for each pixel based on land use in the neighboring pixels, the intensity of the influence decreasing with distance. The authors used this model to simulate the expansion of construction on the outskirts of Bogota.

Hélène Mathian and Marie Piron (Chapter 2) have also focused on the formalization of the consequences of proximity relations. These relations are the result of similarities or, on the contrary, of oppositions between the locations, and the associated spatial structuralization can be determined from local statistics. Such influences can work according to a logic of location pairs (what happens in location A depends on the characteristics of location B) or according to a logic of *context* (what happens in location A depends on the general characteristics of the environment where A is located). The authors show how the statistics formalism contributes to describing and sometimes explaining the result of this influence system on the intensities of the spatial differentiations. A mathematical formalization of the same phenomena is given in Chapter 9, focusing more on the dynamic aspects of the interactions. The geographic coordinates of the spatial entities described by the model then play the role of parameters in the space-time system being studied.

Interaction between agents, groups of agents and the environment

Throughout the previous section, we have focused on spatial interactions as interactions between *locations*. This phrase, which relies mostly on an aggregate logic, is used in order to emphasize spatial logics, but clearly these interactions *between the locations* take shape most of the time as exchanges between agents who live and act in these locations: these are exchanges between individuals, or between individuals and groups, or also between individuals and firms or institutions. Chapters 4, 5 and 6 give great importance to individuals and deal with interactions on this level more explicitly. Thus, Einar Holm and Lena Sanders examine the interactions that drive the decision to migrate and which should be integrated into a microsimulation model: interactions of the individual with other members of the household, interactions with the agents comprising his environment (building, neighborhood, city), and finally, the relationship with the locations where he stopped along his own past trajectory (Chapter 6).

Another form of interaction is modeled by several authors: that which exists between individuals or groups and the environment. Most of the time, these interactions occur in a given space, but they are generally not described in terms of *spatial* interactions. These man-environment or society-nature interactions are discussed more explicitly in Chapter 7, where Jean-Pierre Treuil, Christian Mullon, Edith Perrier and Marie Piron explain the use of multi-agent systems to simulate spatialized dynamics, the driving force of change operating mainly through interactions between agents. The entire functional diversity of interactions is considered: solving computer science problems, simulating negotiation and cooperation problems between agents, managing the evolution of artificial worlds comprised of agents and resources that interact.

From the individual to the group, from the pixel to the area, from the element to the aggregate: the question of the level of observation and modeling

There are two major questions in this field: what is the adequate level of observation given the question we wish to answer? How do we go from one level to another, in terms of semantics and methodology? Of course, no chapter can avoid facing the first question, and several of them explicitly deal with the second.

Choosing the level of observation

The first step in many research studies consists of determining the best level of observation to choose, given the objective of the study. How the question is stated can sometimes lead to an obvious answer: if we wish to know what determines individuals in their choice of transportation from their home to their workplace, we have to work on the individual level; the same goes when focusing on the behavior of working parents and how they manage their daily space-time (see Chapter 4); on the other hand, if we study how an administrative or linguistic border can act as a barrier for migrating patterns, an aggregate level is preferable. Other questions lead to a large variety of possible choices. Such is the case when trying, in an operational process, to determine the localization of a public service (see Chapter 3). The question is to understand, given the distribution shape of the population expressing a need, as well as equity and cost criteria, what spatial aggregate level to choose in order to reach optimal localization. For other questions, finally, there can even be some controversy among researchers as to what level of observation and modeling to choose. In different subjects, there is also a debate between the advocates of modeling on the most basic level and those who suggest developing an aggregate model to describe the dynamics of spatial aggregates (see Chapters 6 and 7). In order to understand the dynamics of water run-off, should we track the “drops of water”, or is it enough to formalize the question in terms of flows? In order to describe the dynamics of population in a set of municipalities, should we formalize the migration choices of individuals or attempt to model the intermunicipality regularities in terms of residential appeal? The problem commonly amounts to choosing between the individual level and the spatial aggregate level, but it also arises, sometimes in a more subtle form, when choosing a level of territorial organization over another (see Chapter 2).

Articulating different levels of observation and modeling

In a certain number of cases, the question is not what choice to make, but instead how to simultaneously use several levels in a heuristic perspective. Thus, the very meaning of both the attributes and the questions asked can vary when covering the different levels of the geographic scale. The authors of Chapter 2 give an overview of the different ways to approach several levels of observation in the context of a study. One method can be to compare the distribution of a phenomenon with respect to different segmentations of space, since such a comparison can help determine the scales of the spatial differentiation being studied (of different forms of segregation for example). It is also possible to compare the type of a statistical relation between two phenomena when changing the level of observation. This relation can be stable or instead reveal stark changes. This last case is problematic from a statistical point of view, since the results depend on which level of observation is chosen [FOT 91], but it provides information on how the spatial differentiations are structured, on the scales of operation of the different combinations of factors being studied.

A similar question is discussed in Chapter 9 with the concepts of scale transfer and "perfect aggregation". The author mentions a common practice in research, in plant ecology for example, consisting of developing a model based on data that are valid on fine scale in order to answer questions on a much broader level. He underlines the gap between the level on which a model is valid and the level for which data is available. This type of approach requires specific knowledge of the relationships between scales, both from empirical and theoretical perspectives. The perfect aggregate from a mathematical perspective (implying for example that if the calculated averages on two sets of elements are equal at a given time, they will remain that way) is rarely observed in practice. When modeling the dynamics of a spatial model, the opposite is often assumed, meaning that spatial differentiations act as a driving force in the evolution of a system. The focus is on the emergence of new spatial differentiations and the change from one scale to the other can then no longer be formalized in a simple way. Computer modeling, particularly the use of multi-agent systems, offers a greater flexibility for managing interactions between objects from different levels, a flexibility that comes at the price, however, of greater difficulties for validation (see Chapter 7).

The problem of interlocking scales is also at the core of the fractal approach. This method focuses on the mechanisms that lead to similar structures on different levels of territorial organization. The authors of Chapter 10 present several methods designed to characterize such structures from a geometric point of view and, most importantly, to locate thresholds in space that correspond to the change from one type of structure to another (a useful method for determining the lines separating different types of peri-urban built-up areas, for example). The fractal approach also usually makes it possible to find the connection between the behavior of a non-linear dynamic system and the type of objects produced by this system. The different steps presented in this chapter are therefore also very useful for characterizing and comparing the structures simulated using dynamic models with the spatial structures that are observed.

Another approach consists of simultaneously formalizing, within a same process, mechanisms related to different scales. They can be integrated into one same model, with for example migration choices formalized on the level of individuals, and residential appeal formalized on the level of municipalities (see Chapter 6). It is also possible to pair up models operating in parallel, for example, by simulating on one hand the potential for change in land occupation in one cell based on the neighborhood, and on the other hand the preferences of individuals for residential choice (see Chapter 8). In each case, the difficulty lies in having to make hypotheses on the type of the feedback between the two levels that are considered.

Multi-scale approaches and the modeling of spatial interactions thus constitute the major themes of this book. Although they deal with different topics and objects, the authors all deal with questions about the organization of space and its role in the dynamics of the phenomena that are studied. The diversity of backgrounds among the authors and of the ways of approaching the problem of space gives an idea of how broad the field of spatial analysis is, and shows the advantage in this field of an interdisciplinary perspective.

Bibliography

[BEN 04] BENENSON I., TORRENS P.M., *Geosimulation: Automata-based Modeling of Urban Phenomena*, Wiley, 2004.

[BRU 92] BRUNET R., FERRAS R., THERY H., *Les mots de la géographie*, RECLUS-La documentation Française, Collection Dynamiques du territoire, Montpellier-Paris, 1992.

[CHA 95] CHARRE J., *Statistique et territoire*, RECLUS, Collection espaces mode d'emploi, Montpellier, 1995

[CHE 95] CHEYLAN J.-P., LARDON S., MATHIAN H., SANDERS L., "Les problématiques de l'espace et du temps et les SIG", *Revue Internationale de Géomatique*, vol. 4, no. 3-4, p. 287-305, 1995.

[DAU 95] DAUPHINE A., *Chaos, fractales et dynamiques en géographie*, RECLUS, Espaces modes d'emploi, Montpellier, 1995.

[FOT 91] FOTHERINGHAM A.S., WONG D.W.S., "The modifiable areal unit problem in multivariate statistical analysis", *Environment and Planning A*, p. 1025-1044, 1991.

[HAG 65] HAGGETT P., *Locational Analysis in Modern Geography*, Arnold, London, 1965.

[LON 96] LONGLEY P., BATTY M., "Analysis, modelling, forecasting and GIS technology", in Longley P., Batty M. (eds.), *Spatial Analysis: Modelling in a GIS Environment*, p. 1-15, Wiley, 1996.

[MAG 06] MAGUIRE D., BATTY M., GOODCHILD M., *GIS, Spatial Analysis and Modeling*, ESRI Press, 2006

[POR 95] PORTUGALI J., "Towards cognitive approach to urban dynamics", *Environment and Planning B: Planning and Design*, 31(4), p. 589-613, 2004.

[PUM 97] PUMAIN D., SAINT-JULIEN T., *Analyse spatiale (1)*, Armand Colin, Collection Cursus, Paris, 1997.

[WHI 97] WHITE R., ENGELEN G., "Cellular automata as the basis of integrated dynamic regional modeling", *Environment and Planning B*, 24, p. 235-246, 1997.

Chapter 1

Modeling Concepts Used in Spatial Analysis

1.1. Introduction

As with many other disciplines, those focusing on spatial analysis have to deal with a large number of complex issues with many particularities. They must therefore try to simplify, in order for the essential detail to stand out, the permanent and determining factor of the contingent and explain the problems in relation to what has been observed. As is the case with most soft sciences, as well as some material sciences, it is difficult for spatial analysis to rely on experience say from physicists or biologists. These are two good reasons that justify the need for simplification and for tools that will help in finding substitutions to traditional experimentation methods.

These two tasks can be achieved with the use of models. Based on a traditional definition by Haggett [HAG 65] and slightly modified, a model is a schematic representation of reality, developed with the goal of understanding and explaining it. The goal has two functions, didactic and heuristic.

To a large degree, we can say that modeling in spatial analysis has come with types of thought that are unwelcome to certain social practitioners or even with a relatively new paradigm. Novelty, which explains a large lack of understanding, is of such magnitude that we felt it justified starting this book with somewhat of a general overview of modeling. On the other hand, modeling, its characteristics and its methods are influenced by the nature of the reality that it studies and by the traditions and the epistemological foundations of the disciplines which practice

---

Chapter written by François DURAND-DASTÈS.

modeling; thus, it varies from one to the other. It is therefore interesting to see how and in what measure the room given to space and the specific aims of the models impose their mark on modeling.

After a necessary reminder of a few logical basics of the modeling procedure, we will attempt to see if the fact that we treat space as a privileged object gives specific aspects to this procedure. We will then focus on the diversity of models used in spatial analysis by proposing a study grid based on a set of binary oppositions.

1.2. Modeling universals

1.2.1. *Logical frames for modeling*

The use of models introduces in our thinking, in spatial analysis as elsewhere, the possibility of using hypothetico-deductive methods; these methods imply an operation following a traditional process:

– reality monitoring, comparisons and analogies combined with *a priori* logical reasoning make it possible to formulate assumptions concerning the operation of a geographical reality;

– this set of assumptions can then be formalized in a model whose first step is purely deductive;

– in many cases, the model can help put simulations in place which will constitute the second phase of deduction. These simulations can incorporate data generally taken from observations; they lead to concrete results that show what reality would look like if it operated in the same way as the basic assumptions of the model;

– the comparison of these results of the simulation with empirical observations will test the validity of these assumptions: a "sufficient" degree of resemblance between simulation and observation can be considered a confirmation, with a degree of probability largely depending on the evaluation criteria of the significance of adjustment between simulation and reality;

– the same comparison will generally highlight the gaps between simulation and observation. We will often refer to these gaps as "residuals", since they represent what is "left" to explain once we have established the explanation by the model's assumptions. The alterations that we can make on the model lead to improvements or deteriorations of the adequacy between simulation and reality; they then become a wealth of information, not only for the content of the assumptions but also for the way the model has implemented them. This information is what is most comparable in social sciences to the contribution that experimentation brings to a large part of material sciences. An inductive approach then brings a complement to the deductive phases of the reasoning;

– alterations are stopped on the model when it becomes impossible for them to improve the adequacy between simulation and observations. The persisting gaps can then be considered "incompressible", at least in the setting that we have chosen.

The study of these "final residuals" is likely to reopen discussions. It can suggest using new models or show characteristics due to facts that were not taken into consideration during modeling; often they could not be taken into consideration because of lack of means or data to approach them, or because of their logical structure.

It may be useful at this stage to explain the process of a very traditional model in spatial analysis called the "gravity" model [REI 31, RAV 85]. Simple in its basic form, it plays a very important part in other forms when incorporated in more complex models. The goal here is to compare the different phases of the hypothetico-deductive reasoning described above in general terms. For practical reasons, we will deviate from the logical order of things and start with phase 2, which is that of model formulation.

Let us take the example of a model designed to show flow relationships in a set of places. These areas are taken into account by pairs. The model has only one equation:

$$I_{i,j} = k \cdot \frac{P_i \cdot P_j}{d_{i,j}^n} \quad [1.1]$$

Its assumption is that a measure of the intensity of relationships between *i* and *j* areas is proportional to the product of the populations concerned by the exchanges and inversely proportional to a function of their distance. The equation then includes a variable that is to be explained, the interaction or the flow $I_{i,j}$, three explanatory variables, populations P_i and P_j from places *i* and *j*, and the distance d_{ij} which separates them. It also includes parameters *k* and *n*. The first parameter shows a general "propensity" to mobility and the second parameter shows the intensity of the distance effect. This is often called "friction of distance".

The development of this model (phase 1) relies on a combination of *a priori* arguments, analogies and observations.

A priori arguments are in part very simple when we consider that the assumption that flows between areas are denser since the areas are more populated and less distant is simply common sense. The simultaneous evaluation of populations by their product and not their sum is less obvious *a priori*. The choice refers to a somewhat abstract notion which comes from combinatory logic.

The choice of the distance function comes from an accumulation of space observations which clearly show that the effect of distance on relationships is not linear. Here, non-linearity has been interpreted by the use of a negative power function (we find decreasing exponential functions in certain formalizations).

The equation of the function is a strong reminder of Newton's law of gravitation. We can say that the model expresses the assumption that the connection between areas is governed by a relation of the same type as that which rules the movements of the planets. In any case, the model owes its name to this analogy.

Based on their orientation, or how much emphasis they place on history, the authors presenting the gravity model [PIN 95] generally focus on the importance of the three elements discussed earlier: observation, reasoning, analogy. It seems difficult for us to completely exclude one or the other of these components from the genesis process of the model.

The simulation (third phase) consists of calculating the value of all I_{ij} by adding in the equation the observed P_i and d_{ij} values for a group of chosen areas. The simulation result is presented in the form of a square matrix; i to j flows are equal to j to i flows, making the matrix symmetrical. Since flow calculation is impossible if the distance is zero, with a power equation there are no values in the diagonal cells.

It then becomes possible (fourth phase) to compare simulation results, by using a second series of data, that of observed flows. It is presented in the form of a matrix that has the same structure than the one produced by simulation; however, it can easily not be symmetrical. The cell by cell subtraction of matrices will show the residuals for each flow and the sum of squares of residuals is an indicator of the quality of the adjustment from simulation to observation. Based on this quality, we can conclude or not conclude to the validation of the model's assumptions, i.e. in the area considered, an important part of the exchange measures is based on a measure of the area's mass and a measure of their distance.

In the fifth phase, we will attempt to reduce the gaps between simulation and observation. The experiment can be done on parameters, particularly on n which, as we have mentioned, measures the deterrent effect of distance. It is also possible to work on variables. For example, for the modeling of certain flows, it might be preferable to measure masses by taking into consideration tertiary assets instead of total population. Similarly, we can wonder which is the best distance to take into account, successively trying distances in kilometers in a straight line, other topographical distances, time-distances or cost-distances. The improvement or deterioration of the quality of adjustment will enable us to draw conclusions of how space works.

The sixth phase, which is hard to avoid, is the one that studies the incompressible residuals meaning here the observed flows that are higher or lower than the flows calculated by the best simulation. This study can be based on a mapping of positive or negative residual flows, as cards will highlight privileged relations or barrier effects. These must then be submitted to an explanatory examination, that will often be based on the researcher's culture and therefore on facts outside of the basic model.

It is worth noting that these maps of residuals based on the gravity model are one of the ways that we can use to represent our knowledge about flows. Indeed, despite generally successful heroic attempts, it is almost impossible to show all the flows on a map if the number of starting and ending regions is higher than 10. We must therefore choose among the flows; the solution often used which consists of keeping the most important ones is limited. It may be much more useful to keep the major residual flows, which give more information on the characteristics of the space. The approach consists of using the model to control the trivial effects (those of masses and distances), the residuals highlighting specific effects which in other circumstances would remain perfectly hidden by trivial effects.

The previous description of the hypothetico-deductive path which contains the modeling practice and the gravity model examples implies that all the conditions for a complete and precise implementation of the procedure are present. Often this implementation must be adapted to take into account concrete conditions of the scientific study and to reach significant results, despite problems with formalization or lack of data. Simulation precision is the major problem encountered. We are examining the case where the nature of the model and the data supplying it enable a precise simulation of an exhaustive set of individual cases and where other data, just as precise and exhaustive, will enable a complete comparison of the simulation and observations. We can say that we are able to make an *individual/individual* comparison between simulation and observation. However, it might be impossible to obtain such precision, either because of lack of data or because of the model's characteristics. We must then test the model's capacity not to simulate a given distribution in a complete form, but a distribution whose structure is the same as a situation that is observed. We can then say that we have a *structure/structure* comparison and not *individual/individual.*

For example, certain diffusion models [SAI 95] simulate points distributions in space or point patterns. It is not necessary for the validation of model assumptions, that this simulated pattern coincides point by point with an observed pattern; it may be enough to compare certain geometric pattern properties, for example, density, frequency, or degree of regularity or concentration of the distribution of points – all quantifiable properties, which makes structure comparison possible. In this case, the nature of the model itself makes *individual/individual* comparison impossible and useless.

1.2.2. *The language of models*

As previously defined, models are intellectual constructions that represent reality by oversimplifying *and therefore* by distorting it. This distortion which is a required condition of their use necessarily operates a selection among the components of the empirical reality. If we stick with the term's etymology, it is obviously an operation of *abstraction*. Models can have different shapes, in other words, they can be expressed in different languages. With regard to them, it is important not to make a mistake in the handling of the notion of abstraction. As an operation, it is essential to define the modeling approach, but it does not necessarily result in an expression that will only handle symbols that have nothing to do with everyday language. There is no simple mandatory link between the approach where the model was developed and the form that it takes. Hence, a mathematical formula, which is a very abstract language, can sometimes be the result of a low degree of selection in a set of empirical observations, with very little resort to thought or analogy and therefore a moderated abstraction study, whereas a drawing – a more concrete language – can be the product of very elaborate thinking.

Even if we keep in mind this necessary precaution, it remains useful to classify modeling languages by degrees of increasing symbolization, or if we prefer, increasing abstraction.

1.2.2.1. *Material or physical model languages*

Material or physical model languages [DAU 87], visible reproductions of real objects, are part of modeling, insofar as the transformations made make them handleable and enable simulations. The most commonly used operation is based on scaling, the production of "scale models", but it is not the only one.

The interest for physical modeling is obvious in the case of "arts", engineering as well as architectural – wind tunnel testing of an airplane model makes it possible to make appropriate and timely choices. A lot of these physical models are really spatial arrangements, such as, for example, the model for a neighborhood or a port where we want to study the effects of a new dam on waves and currents; this would suggest that the use of physical models could bring enormously useful information for our knowledge basis with spatial analysis. It is unfortunate that this method is rarely used by geographers, at least as far as the heuristic aspect of modeling is concerned. It would seem, however, that it is different on the didactic side, where physical models have been used for several demonstrations. The physical formalization of the model of industrial location by Weber [HAM 67, WEB 09] seems to have mostly didactic virtues.

1.2.2.2. *The language of images: iconic models*

Iconic models attract attention, on the one hand because they are widely used and on the other hand because they are at the center of much heated discussion.

It is understandable that cartographic images are the first to be noticed. We can say that each map sorts between different aspects of reality, so that it gives a representation of reality that is necessarily altered, even if only because of the double constraint of scaling and of the transition from sphere to plan. Many maps, even the most widely used ones, are based on approaches similar to what is the basis for modeling. For example, the one that consists of using a relation to eliminate trivial effects in order to emphasize important specificities. In this way, isobaric maps currently used are based on a calculation of observation correction. The fact that it was referred to by a seemingly meaningless expression, “mean sea level”, and that the community of users is very familiar with it must not overshadow the fact that this is the application of a physical atmospheric model, the standard atmospheric model, and that these maps have reached a high degree of abstraction.

In this perspective, we can envisage a hierarchy where the cartographic images are arranged according to the author’s will for conscious schematization, to the part he assigns in the map building process, to simple registration, to inventory, or to the willingness to explain, compare and demonstrate.

Used recently, mostly in France [BRU 87], very abstract cartographic schematizations follow two different routes, where inductive and deductive processes intervene at different steps of the study on reality (see Figure 1.1).

A first process starts with a complex reality, made up of complex information on different empirical cases, often presented in a cartographic form. A simplification (and therefore an abstraction) from these data leads to the development of a “schematic map”. It can be considered as a model if comparisons and generalizations can be achieved, as well as synthetic summary formulations of a set of spatial processes. Practical examples of this method are numerous; we will find one particularly significant application in “Lieux de mémoire” by Grataloup [GRA 96], who uses them to bring out what he calls the “principles” and that we could also consider as models of spatio-temporal organizations.

The second approach starts with elementary spatial models, or with products from the analysis of more complex models, and brings out from these arguments a few elementary spatial forms, which can be called "choremes". The credit for formulating the term and developing the method goes to Brunet [BRU 80]: it is the result of the combination of the radical *choros* (place, space) with a suffix showing the relationship with the "phonemes" used by linguists. The goal is to show that the combinations of a limited number of choremes can produce a large variety of spatial situations, just as phonemes produce the infinite richness of a language. The combinations of choremes can be used to describe unique spatial configurations, which are thus related to the rules taken into account in the starting models – particularities are therefore interpreted according to general principles. We can also build recurring combinations of choremes, the "chrorotypes" that can be useful in describing and explaining spatial organizations.

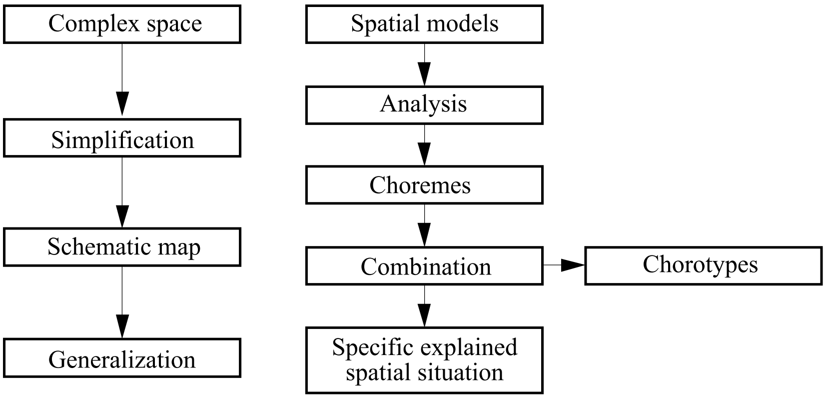


Figure 1.1. *The two possible approaches to cartographic schematization*

It seems that both approaches have significant advantages, but that a certain confusion has taken place between the two procedures and that it has sometimes be thought that simplifications with simple drawing techniques would be sufficient to define a "chorematic" approach.

A second series of images for the demonstration of models is formed by what is conveniently called sagittal diagrams. They are boxes containing elements of the studied set and arrows indicating a relation between these elements – hence their name. It is then possible to symbolize several types of relations:

– relations linked by going from one step to another during the progression of an intellectual process. It is the meaning given to the arrows in Figure 1.1; this symbolization is general enough for someone wishing to develop a logical progression or explain it to a reader. The same logic applies to computer scientists' flowcharts which precede or follow application development, diagrams that show microsimulation procedures and many more diagrams;

– cause and effect relations. In general, we agree that the element at the base of the arrow is the cause of the element placed on the side of the tip. The interaction relation can be equally demonstrated by a double arrow or two arrows. It could be useful to insert in an intermediate position a logical element in the causal relation, by using of an explanatory written form in the legend;

– more concrete relations such as, for example, flows between elements that are themselves more concrete, specific portions of space (regions *i*, *j*, *k*, etc.) or types of space (cities and villages, rural spaces and urban spaces, etc.). The flows can obviously be tangible and quantified (number of tons, number of trains or planes between *i* and *j*), or intangible and quantified (number of bits of information passing between *i* and *j*) or even tangible or intangible and not quantified (there is a relation between *i* and *j* in commercial exchanges).

These flow diagrams, which are more abstract and thus more manageable than maps, currently constitute spatial models even more so than those based on cause and effect relations. These diagrams are sometimes considered as premodels or models of models; we could potentially call them *metaspatial models*, which justifies their mention here.

This seems to show clearly that, as with maps, sagittal diagrams play different roles in modeling, albeit always important ones. We increase their readability considerably and their usefulness if we accept to modify the figurative signs used for the arrows when they represent different relations in a diagram and of course when we accompany this with a legend.

If the maps, cartographic schematizations and sagittal diagrams have become very important in terms of modeling, especially in current practices; we should not forget the existence of a large variety of images, which are sufficiently elaborate and oriented toward simulation to be considered as models. To only mention one case, despite the reluctance proclaimed for modeling by the authors of a number of French textbooks of physical geography, we must recognize that geomorphologists, for example, often present relief evolution models in the form of iconic diagrams, such as geological profiles or block diagrams.

1.2.2.3. *Modeling in mathematical language*

Modeling in mathematical language has assuredly reached a higher degree of abstraction, at least from the point of view of language, which is not the only one to take into consideration as we have previously mentioned.

Mathematical language (see Chapter 9), or in other words highly symbolic formalization, is in fact used for implementing order in the empirical reality. This order building can be considered as part of modeling, although using very little *a priori* reflection and analogy at the start of the method implementation [CHA 95]. Representing a group of spatial individuals described by two variables from a scatter plot and summarizing by calculating an adjustment curve is definitely part of modeling. It is in fact the development of the representation of reality and the curve provides a simulation which, when compared with observed data, enables us to calculate residuals. The starting hypothesis that will be tested might seem limited: it is of the "there is a relation" type. This hypothesis is limited because it is the only one, but not really weak since it is full of implications. The procedure's aim is to validate it. This latter statement explains what we consider the existence of statistical modeling, which we also call *descriptive mathematical modeling* for reasons which will be explained later.

A second aspect of modeling based on mathematical formalities consists of using it to describe and quantify relations chosen according to a series of *a priori* assumptions. These assumptions come from basic reflections, analogies and assumptions from other disciplines. That is how several models of urban space have been developed from the "economic base theory" combined with gravity, with more general implications because the link is obvious with the design of a *homo œconomicus*, which is the basis of reasoning for what we can call traditional economy [DER 79, LOW 64]. In a kind of phenomena that is quite different but also related to spatial analysis, the basic assumptions of fluid physics have been used to model atmospheric circulation, a spatial analysis of air stream at different levels.

A strongly established tradition consists of building these models in the form of a set of relations interpreted by several complex differential equations. They are often linked in order to form systems [SAN 92, WIL 74].

Relations modeled by these equations mainly calculate flows and thus stock variations, either directly or by totaling flows; we can find the stock's total value at a given moment. This is a very distinctive advantage of mathematical modeling, insofar as it enables us to use the method to make predictions. By verifying the consistency of equations and by calibrating parameters with known data, we could risk predictions by extrapolation. This advantage is also a disadvantage, since spatial analysis cannot exceed its purely operational frame in order to become a heuristic value without taking into account any other relation than the flows.

The consideration of both flows and logical rules seems to be one of the most interesting characteristics of what we could call cybernetic and computing techniques (see Chapter 7). These techniques rely less on the use of equations than on rules that control the transformation of "individuals", such as cells or groups of cells in a matrix. What mainly concerns us is the possibility of introducing rules which explain the evolution of an individual from the evolution of other individuals in relation to it in space. Relative positions are thus introduced as explanatory elements which are essential in any spatial analysis. This type of modeling can take many forms: multi-agent systems, cellular automata, various forms of microsimulation, etc. The major design differences must not allow us to forget that there is a fundamental relation between these techniques [BON 01, BUR 96].

There are common traits between some of these computer modelings and the use of physical models. In fact, each cell can be equipped with reactive properties that we find in physical models, linked to concrete representations of real objects. The cells, like the elements of a "scale model", are then manipulated by the modeler. On the other hand, computer modeling, which is more abstract, is more flexible. That is why it tends to replace the construction of material models in the studies made by engineers, in the study of interactions between forms and flows, whether it is in architecture, hydraulics or aeronautics.

1.3. A few specific features of spatial models

To have as a basis declared methodological principles and the use of various languages is a characteristic common to all disciplines that use models. Spatial analysis is therefore part of a general movement when it implements them; it is not alone in this perspective. However, if radical specificities in terms of modeling do not come from characteristics inherent to spatial analysis, which would give it a place all of its own, we can still notice that the discipline has specific characteristics in the practice of modeling. They are mainly the product of constraints that give much weight in modeling to spatial characteristics, whose introduction as a priority is an obvious necessity in the matter.

The introduction of space as an essential component of models has significant consequences on certain characteristics of their structure. They must in fact be able to implement a fundamental principle, spatial interaction, which says "everything that is located or happens in a given place is at least partly determined by all that is located or happens in a group of places in relation to the considered place" [DUR 90]. These relations between places, in terms of reciprocal influence, must be formalized. They are often formalized in an implicit way, although some do not find them sufficiently clear in iconic models and especially in cartographic diagrams.

In his well known model on French space, Brunet [BRU 80] implicitly makes the assumption that the interaction between the French axis and the European London-Milano axis largely explains the fundamental asymmetry (at the time) of the French territory. The multiple processes of this interaction remain in a "black box". The matrix language adds more precision. The basic matrices are square matrices, where analyzed places are depicted in rows and columns, very often as “origin” in one position and “destination” in another. The matrix cells can include qualitative and quantitative information:

– qualitative such as, for example, “there is a relation”, “there is a flow”, “there is domination of the origin over destination”, etc. It is probably not necessary to mention that these qualitative data are generally conveyed in a numeric code;

– quantitative, insofar as we can include measurable data. These data are most often either distances or flows.

One of the traditional practices of spatial modeling consists of comparing flow and relation matrices with distance matrices, whose importance translates that of relative positions in the reflection on space and on its role in the operation of human societies. This comparison can take the form of various calculations, with purposes that are different from one model to another. It can be useful to state that the matrix distances used here can be very different from those used in mathematics. It is in fact possible to consider that certain distances can be asymmetric, especially those represented in terms of price or time. Furthermore, if we consider paths between areas and not points, we can state that the distance between a spatial object and itself is non-zero, which is unacceptable for the mathematician. As a consequence, purists gladly reserve the notation d_{ij} for the mathematical distance and another, for example c_{ij}, for the less orthodox distances.

Sometimes relative positions are given by distance with regard to one of several starting places preferred because of their singular qualities, such as the operating basis in Von Thünen’s model or the group of models generally derived from it [HEN 67, VON 27], or like the downtown part of a city in several urban space models. Generally, there is a set of models built on the opposition between “central” and “peripheral” spaces [REY 81], which will take into account one or several points or areas used as the origin of the distance to be measured.

The importance of spatial interaction is at the heart of a logical chain that explains a very important characteristic of spatial analysis in our view. Spatial analysis logically leads to an important place given to all forms of interaction, to retroaction loops as well as to reciprocal causalities; the role of the latter leads to the use of the frame of systematism as a formalization mode [DUR 81, DUR 84]. Therefore, it seems normal that the system models occupy such a large place among spatial analysis models and that this important place is linked not only to the importance of systematism in the contemporary line of thought but also, and especially, to the non-linearity of causalities which take place in spatial relations – which sets them apart from temporal relations who are subjected to the constraints of the “swift passage of time”.

There are two types of systems implemented in spatial analysis, as we have previously implied. In the strictest sense of the term, spatial systems contain spatial entities as elements and the relations are material flows, or actions located in space. However, we can also use logical systems, in which the elements are abstractions, notions or general propositions and the relations are cause and effect links.

Spatial systems can be formalized as sagittal diagrams or, as is often seen, as mathematical forms which are more rigorous and will enable calculations.

By applying the logic of the prey/predator model to urban growth, Dendrinos and Mullaly [DEN 83] have proposed a model for urban growth based on the simple assumption that the difference between average income per capita in a given city in a space of reference, for example a country, and all average incomes per capita for all cities of that country will determine migration toward the cities where these differences are positive. This migration then explains population variations and therefore the total population at the end of a given time. According to the assumption of the model, since each city has a production potential that cannot support population above a certain limit, when it approaches this limit, population growth reduces income per capita. The systemic character of the model is obvious because for each city there is interaction between the income per capita and the size of its population. Additional elements are introduced in the operation of this initial loop: migrations toward a city are limited or increased, all things being equal in terms of income, by the intervention of diseconomies or economies of scale. Furthermore, different populations will react differently to the variations of incomes.

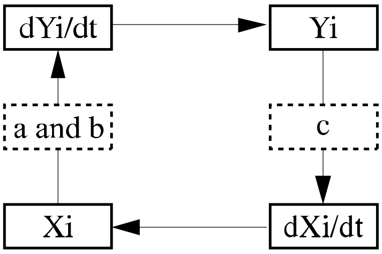
Mathematical formalization of the model rests on two equations which make up the system; the variables are the numbers of populations and incomes: variations of the former are caused by demographic migrations and those of the latter by demographic migrations, whereas the scale effects and reaction speeds are taken into account by the parameters of the equations. We then have:

$$\frac{dX_i}{dt} = a.(Y_i - \overline{Y}).X_i - bX_i^2 \qquad [1.2]$$

$$\frac{dY_i}{dt} = c.(X_i - X_i^*).Y_i \qquad [1.3]$$

where X_i and Y_i indicate population and income per capita respectively in city i, $\overline{Y}$ indicates the average income per capita in the group of cities in the country and X_i* the limit of the population of i according to its production potential.

Parameters *a* and *c* refer to reaction speeds and *b* measures the economies/diseconomies of scale. If there are diseconomies, then *b* is positive. It becomes negative in the case of economies of scale. Population growth is slowed down in the first case and accelerated in the second by the growth itself. The square term of the first equation brings in a limiting factor.



Interactions between state variables and the variations of these variables

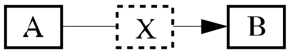
N.B. The diagram  means: "A produces B through X" but not "A produce X which produces B".

Figure 1.2. *Urban growth system, mathematical language*

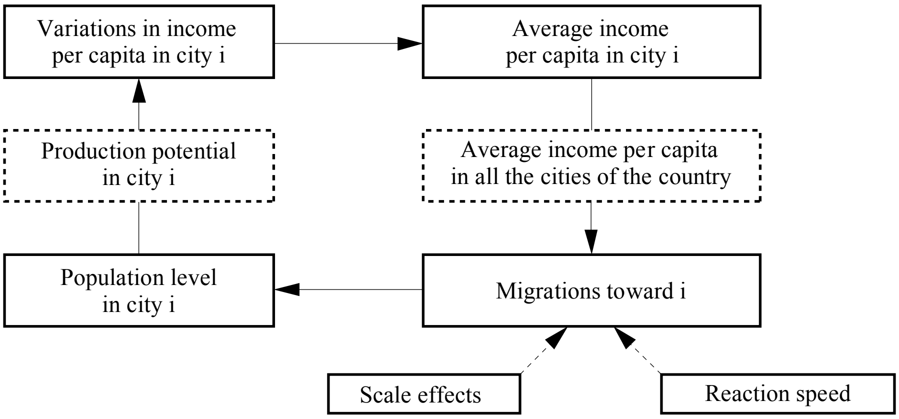


Figure 1.3. *Urban growth system, regular language*

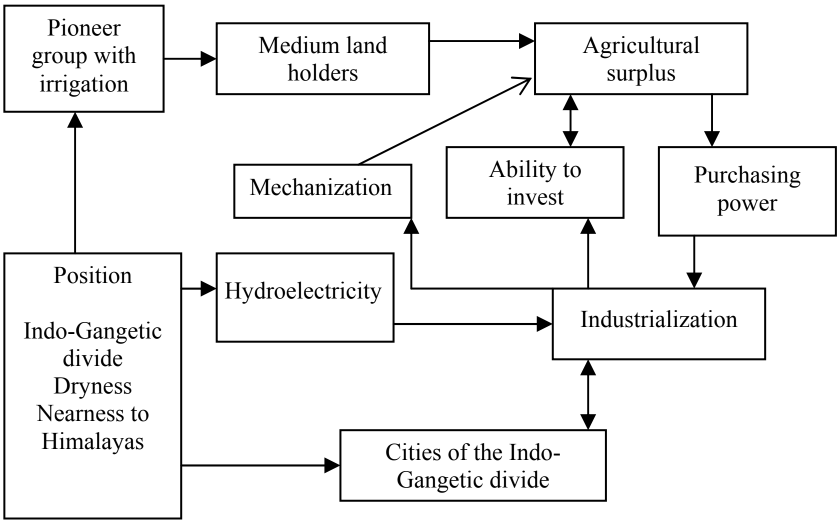


Figure 1.4. *Pendjab system (according to Géographie Universelle Reclus [DUR 95b])*

It is clear that the equations form the system: values calculated in one are used in the calculation of the other. This systemic character can be highlighted by a sagittal diagram, or even by two diagrams of this type: one using the same formalism than the equations and the other written in regular language (see Figures 1.2 and 1.3).

There are numerous cases where it is interesting to really understand the multicausal and interactive character of the spatial analysis explanation to use sagittal diagrams where the relations do not lend themselves to mathematical formalizations, either because they are not quantifiable on account of lack of means, or because they are too complex to be included in mathematical formalizations. We can only hope that examples such as the one we are proposing concerning the uniqueness of a region, the Pendjab located in the northern part of India and known for its economic progress compared to the rest of the country, will show the usefulness of this type of graphical modeling (see Figure 1.4). It seems to us that their use could avoid numerous errors and shortcomings in the explanation of spatial facts mainly because of the difficulty of breaking with mental habits revolving around linear causality.

1.4. Spatial models: a study grid

The practice of modeling in spatial analysis offers a large range of possibilities of studied reality and available technical possibilities, according to the goals we are seeking to achieve. These multiple choices can be structured around a series of concepts, each with its own antithetical concept, so that we can present them in binary opposition form. They can then be grouped again to produce a classification tree, which enables us to orient ourselves in the large variety of possible models. Of course there are other methods to classify the models. We have seen the interest and the limits of those that only use the languages. A fundamental study on models in geography [CHO 67] was based on purely operational criteria: hydrologic models, urban models, agricultural models, etc., which is obviously convenient for users but does not encourage general considerations. The binary opposition procedure has the advantage of questioning the notion of explanation of the respective place of individuals and groups, of the determined and random facts, and of the concept of balance.

1.4.1. *Sequencing and explanation*

First, let us recall the types of thought integrated within the design of the models, since they distinguish themselves by the complexity and the level of conceptualization in the incorporated assumptions. There is a lot of ambiguity in many works and linguistic practices, possible debate sources, and it is not a bad idea to explain in more detail the examples of sharply contrasting approaches.

Let us look at the relation between agricultural productivity of the soil ("yield" by surface area) and the density of the agricultural population. We can easily assume that there is a relation between both and we can test it, or we can simply try to put some order in information that is already rich in significance. Researching such an order can be accomplished by drawing scatter plots, by calculating correlation coefficients, regression line equations, etc. All these approaches lead to "schematic representations of reality", "elaborate" and useful in order to understand it. We can thus consider that they come from modeling. If we add that the best adjustment curve is a straight line, we can say that the relation is part of a "linear model". Establishing the fact that the correlation between the two variables is linear, positive and statistically significant, as is often the case, leaves a lot of questions unanswered, mainly the one concerned with the process at the basis of the relation. In the example used here, it is either high demand for consumption due to high demographic density that has mandated the use of techniques ensuring high yields, or the adoption of such techniques that has enabled high population growth by reducing mortality and/or by attracting migrants, or yet again there was loop interaction between the two events. The linear adjustment also highlights important residuals that require other research procedures in order to explain the phenomena.

In short, the model is based on a limited hypothesis, it tries to compare a complex situation with a mathematical order and it is qualified according to the nature of the mathematical structure that summarizes this order (here it is a linear relation). All these characteristics compel us to tie the research procedure to what we call data analysis; but we cannot totally exclude it from modeling, for reasons recalled previously In these cases, we can conveniently talk about "descriptive modeling" while adding that we do not intend the term to have any pejorative connotation.

It is therefore in terms of difference, and not of superiority or inferiority, that we must situate the contrast between descriptive models and those that we refer to as "conceptual models", insofar as they reference conceptually rich backgrounds and are largely based on deductive operations. The very traditional central place model of Christaller illustrates this concept in a highly eloquent way [BER 70, CHR 33, PUM 95]. This model is based on traditional economic theory, insofar as it is rationalized from the behavior of a perfectly rational and perfectly informed "homo œconomicus", concerned with maximizing his profits and minimizing his efforts. It uses the abstract notion of service that it places in a hierarchy according to the associated characters (frequency of recourse, admissible distance, customer threshold) summarized in the concept, itself rather abstract, of centrality. From this economic theory, and particularly from the centrality of the services, we move on to spatial results: centrality of places in a hierarchy with discontinuities, then optimal arrangements helping to maximize the satisfaction of both groups of economic operators, the suppliers and the service consumers. A geometry based on an essential figure, the equilateral triangle (the highly praised hexagon only comes into place later), is inferred from these premises, with some nuances depending whether we want to decrease distances to a maximum or whether we prefer to take profitability of transport networks into account. We should add that the introduction of administrative control constraints of the territory causes huge changes in the organizations when we deal with large areas.

It would be useful to for us to use very conventional data once more, to emphasize the variety and the richness of references to general concepts and theories that are the basis of the model; a richness and a complexity that are sometimes overshadowed in certain presentations. We will point out that this model, which is highly deductive and abstract, is expressed in an iconic language, even if it can also be the subject of mathematical formalizations. We are far from a simple empirical data order of a simple ordering of empirical data, even though the observation of networks of cities was at the basis of the model builder's questioning and that the model has been developed to explain observed realities.

These two examples would then tend to suggest a first binary opposition between descriptive and conceptual models, between models that primarily want to complete sequencing and models that give a large place to theoretical and conceptual references. However, the association of these two types of intellectual operations is frequent: there are a good deal of *a priori* notions in the analysis of data and a lot of empirical references in the deductive and conceptual progressions, so much so that between models it is more a question of gradation than of rigorous opposition. That is one of the reasons why we shall keep this first distinction in the background, without integrating it in the final classification of models that we shall retain. It is still important to emphasize the fact that in numerous studies using models, these will provide a necessary ordering which will help other types of arguments likely to deepen the explanatory procedure.

1.4.2. *The group and the individual*

In spatial analysis as in other disciplines, the models are often aggregated; in other words, they process group behaviors or average behaviors from a multitude of individual behaviors converging to produce global results. In general, it is the parameters that take into account the description of these collective behaviors. In the case of equation 1 of the Dendrinos/Mullaly model (see equation [1.2]), parameter "a" has a relatively high value if the population reacts swiftly and/or strongly to differences in income, or low in the opposite case. If the population was in general indifferent to the differences in income, the parameter would be zero. From a heuristic standpoint, at least ideally, the course followed is inductive: the value of the parameters that produce the best adjustment between simulation and observations enables us to learn about the type of behavior of the population, its sensitivity to differences in income in our example. What we are doing here is experimentation from models.

These aggregated models have been extremely useful; they also have their epistemological justification. However, they have also aroused dissatisfactions because of the problems that the notion of group behavior entails, or even "average" behavior. Dissatisfaction is vigorously expressed by geographers who insist on the importance of actor decisions, of their subjectivities, and claim to adhere to "social" or "humanistic" geography.

For a long time, economists have tried to use as a starting point individual characters of economic operators by implementing functions of utility and choice models.

These are based on the confrontation of several sets: a set of individuals, a set of characteristics or attributes of these individuals, a set of possible choices and a set of attributes for these choices. Considering the attributes of individuals and choices, a certain choice presents a certain usefulness for each individual. By assuming that the individuals try to maximize the usefulness for them to make a given choice, we can define the probability for each of the choices to be made by the set of individuals. For example, knowing for a set of households their status, their income, their composition (age of parents and children, sex, professional activities, etc.) and the characteristics of available housing (rent, cost, location, property, size, collective or individual character, etc.), we shall attempt to calculate the probability of the demand for each combination of attributes, for example: small rental collective apartment, large standalone house, small owned bungalow, etc.

These choice models are inserted in various argument procedures, with goals and possibilities, based on what is known at the start and what is researched as outcome.

A "retrospective" progression starts with the knowledge of attributes of individuals and choices, and the actual choices made; the goal of the research is the determination of the value of the group of parameters measuring the share of each of the combination of attributes in the explanation of these choices (for example, of how much does the fact that the household has more than five members increase the probability that it will choose a bungalow?). These retrospective studies will then allow "prospective" progressions starting with the knowledge of the attributes and of the parameters measuring the respective role of the different attributes in the probabilities of choices made by different types of individuals. Therefore it will be possible to indicate the expected recurrences for each choice. In this way, we go from observed frequencies to probabilities and then from these probabilities to expected frequencies.

This type of modeling, which is part of spatial analysis as long as the choices have a spatial significance, is of obvious prospective interest. It has actually been widely used by Dutch geographers, in a country very concerned by urban planning, which explains the place given to choice studies in housing. Arguments of the same type, based on research of probabilities in associations of attributes, with the same formalization but concentrating on different entities have been used for purely explanatory reasons, or if we prefer more speculative reasons, and in a more directly spatial way. In the same spirit, *logit* or *probit* type models enable us to calculate the measure in which attributes of a given area increase the probability of observing a certain character of that area. For example, in what measure the permanence of occupation of an archaeological site (the variable that we aim to explain) is associated with attributes of that site, such as its proximity to valleys, its topographic location, its altitude, etc. [FAV 98]. The frames are different from the one that is used with choice models, but the logical structures are similar, mainly when it concerns the role played by the consideration of each individual and the role of probabilities. The advantage of these procedures is that they associate variables of different types, qualitative and quantitative, and determines the role of each one, all things being equal for the others.

Individual characters, their behaviors and the spatial consequences of these behaviors are also implemented in microsimulation procedures (see Chapter 6). The model is based on a group of individuals with attributes and rules that control changes in individuals according to their attributes and those of the individuals with which they interact, including spatial interaction. We can then either validate those rules, or use them to predict on site evolutions and movements, and therefore qualitative and quantitative modifications of the population from each area of the space analyzed.

1.4.3. *The random and the determined*

The introduction of the individual dimension in models almost invariably means resorting to the notion of probability. That is a partial aspect of a more general problem, the respective share of deterministic and probabilistic models in spatial analysis.

We know that the deterministic point of view in its exact sense, that of Laplace, is based on the idea that a complete understanding of the state of the universe at time *t* would give a complete prediction of the state of the universe at time *t+1* later than *t* (complete understanding that would only be possible for an omniscient "demon", but that is another story). In modeling terms, the deterministic principle presumes that of the variables, rules and relations of the model we can have one single resulting situation, which is completely determined by these variables, these rules and these relations. Most of the traditional spatial analysis models are deterministic. It is in fact easy to verify that in models such as those of Christaller, Von Thünen or Weber [CHR 33, VON 27, WEB 09], we can only have one single spatial configuration as a result of the variables, parameters and rules introduced in the model.

As useful as these deterministic models may be, the complexity of reality escapes from them sometimes and it is often necessary to introduce random elements, an operation that can take many different forms.

In the first place, randomness is often used as reference to test the significance of observed relations. We already know that inferential statistics are largely based on testing the difference between an observed relation and one that can be modelized by a random process. In spatial analysis, this common procedure can take in certain cases specific aspects, as with the study of point patterns. These point patterns can be described by single numbers (average distance from each point to its closest neighbor, for example) or series of numbers (number of boxes in a grid containing 0, 1, 2, … *n* points). Statistical techniques make it possible to calculate the value of these numbers or the structure of these series where the pattern is "random", i.e. when the presence of a point in a place does not change the probability of the presence of a point close to this place, that is, in the case where there is no spatial interaction or no dominant spatial interaction. We can then compare the observed numbers or the series of numbers with those calculated in the hypothesis of randomness and draw conclusions if the differences are significant.

Secondly, a random element is often introduced in the simulation process itself. We set a certain number of rules and relations, but in the implementation of the simulation and according to these rules, we use a draw. The procedure is particularly used in migration or diffusion models. In diffusion models [SAI 95], the probability for an innovation to reach such spatial individual is set in the model, but the location of this individual during simulation is given by a draw based on these probabilities (see Chapter 5). Random draws are introduced to "show" the intervention of individual characteristics or even to take into account what is left of human liberties considering the constraints (spatial or other), the weight of which the models are attempting to understand.

Microsimulation procedures widely use probabilities and random draws; computers becoming more and more powerful make it easy to generalize the rule-probability-draw association by multiplying the steps and taking into consideration a large number of individual characteristics for a large number of individuals.

It is obvious that in most cases these random simulations attempt to produce not an exact image of reality, but a view of its fundamental structures. The comparison here will be what we have previously called a structure/structure comparison and not a case by case comparison.

1.4.4. *Movement and balance*

Truisms can sometimes be useful. It is certainly a truism to say that if a situation can be observed it is because it has made its appearance, has evolved to take the shape that we are familiar with and, has remained; but this truth is sometimes lost from sight, so much so that we should be reminded that the explanation is incomplete if it does not take into account these different components of the origin of the observable realities. Since they often follow different logic, we do not reach the same level of understanding for each of its components and we tend to focus on one or the other, or even to ignore one of them. It is a legitimate, sometimes inevitable, choice; still it should be made consciously. As with other explanatory approaches, modeling must face this problem.

Many models exist which focus on the way a situation remains steady or even ensures its continuity. We shall see that this is mainly the case with several of the oldest and the most traditional models in spatial analysis. These models are easily described as static, insofar as they attempt to show above all else the mechanisms and conditions maintaining a spatial structure, and they highlight the problem of the relationship of this structure with its environment in a synchronic perspective.

The notion of a static model has a certain number of implications that can sometimes be a source of problems mainly in the definition formulation. Indeed, we often associate the notion of balance to that of static model. This notion of balance is defined *stricto sensu* as a situation of immobility caused by the compensation of opposing forces. However, it is permitted to associate notions of movement, or even of evolution, to that of balance: thus, in atmospheric physics, a wind field, and therefore movements, can result in *geostrophic equilibrium*, which is the result of different forces: Earth's gravity, pressure force, centrifugal force, Coriolis deflection (effect of the Earth's rotation that may conveniently be introduced as a force). The wind field is considered as balanced because among other reasons, the different forces combine and compensate for each other and maintain the structure of this field, the essential features of its form. The cases of maintaining the form and the structure, while the elements change shape, are found in numerous other domains. That is how, for example, we can say that an urban network is balanced where relations between the size of the units stay the same, despite the variation of this size; in other words, the curve that relates the sizes and the ranks keeps the same form, while going through a translation. This continuity of the form can be interpreted in terms of balance. In order to take what has been said into consideration, it can be useful to make a distinction between fixed and mobile balance, even if such notions can open a debate.

The necessity to set apart stable balance, i.e. able to recover when affected by disruption, and unstable balance where this recovery does not happen, should not be too problematic. There is, however, an unfortunate tendency to confuse balance with stability, despite the importance of unstable balance.

It would be pretty easy to verify that a good number of traditional spatial models are based on maintaining structures by the play of contradictory actions which are able to resist disruptions. Christaller, Weber and Von Thünen did not reason any different or on other bases and there would be many other examples that we could mention from more recent models.

Static models have been very useful, in spatial analysis as in other domains, from physics to economy. But they have disadvantages. Not so much the one which we would think obvious, excluding time: time can be reintroduced with the help of diachronic "compared statics". By analogy, the comparison between photographs of the same place at different moments is much more interesting than one single photograph (even if the cinematography alone can reveal the actual dynamic).

The most disturbing thing with static models is that they neglect large sections of the explanatory issues. They only make sometimes implicit weak assumptions on emerging mechanisms of modelized situations. In the case of traditional models mentioned previously, the only procedure that seems to have been considered is trial/error and, furthermore, it has not been clarified. We can consider, for example, that if a group of service contractors sets up, for an unfortunate reason, in a "badly located" area in relation to the Christallerian network, they will pay for that *error* with a failure and will need to give up. This is happening as if a move from an equilibrium led to a return to the starting position. As an example, the city of Richelieu, which was built according to the order of the illustrious cardinal minister in the South of Loire, went back to being a small unimportant town when he died and ceased to put pressure on the local elites to maintain houses in the city. As a result, you can today admire Louis XIII architecture towering over sunflower fields.

With regard to balance situations, static models do not say how they have been reached. A shortfall in itself, but made worse by the fact that we are unable to ask some important questions: were there other possible balance situations, and if so, why have we reached this balance and not another one? How and on what conditions can there be observable situations (permanent enough to be observable), that do not correspond to a balance [PUM 89]? And still: since situations characterized by rapid fluctuations are observed or definable in theory, how and on what conditions are they reached or avoided? We can see that there is a whole series of questions opening onto the notions of order and chaos, on which we cannot dig any deeper here [DUR 91].

Let us simply say that a family of models, dynamic models, looks for answers to the questions that were expressed above. They most often take the form of differential equations or of systems of such equations; the latter make it possible to reproduce evolutions and explain them, consider different possible evolution types and attempt to understand why we are observing in fact one instead of the other. Even simple equations provide rich query possibilities, as an example will show, which is convenient because of its simplicity.

Rogerson [ROG 85] considers the size of store W_j (for example, the sales surface), the income D_j provided by sales, which can be formulated in terms of client demand, k_j unit cost of operation (all costs brought to the sales surface). An elementary dynamic model of the evolution of the size of the store is:

$$\frac{dW_j}{dt} = a.(D_j - k_j W_j)W_j \qquad [1.4]$$

where parameter *a* measures the speed of adjustment or, if we prefer, the sensitivity of the businessman to loss or benefits.

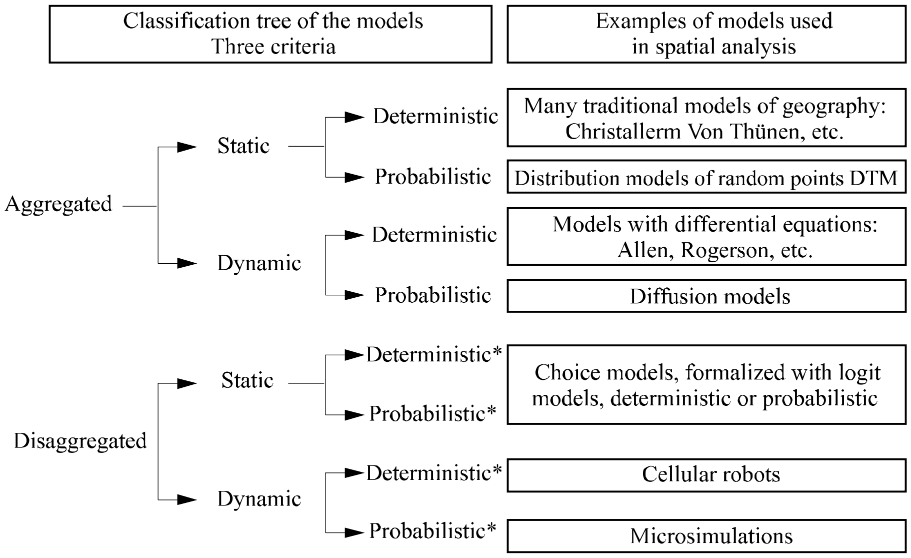
The writing of differential equation in the form of equation of difference:

$$W_{j(t+1)} = W_{j(t)} + a.(D_j - k_j W_{j(t)}).W_{j(t)} \qquad [1.5]$$

enables a series of iterative calculations that we can make by changing the initial conditions (value of starting variables) and parameter *a*. This experiment produces interesting results. The simulation produces completely different results, based on the combinations of chosen values, mainly following the relation between the initial value W_j ($W_{j(to)}$) and of the quotient $(1+a.D_j)/a.k_j$. Based on the results, either the store is not viable (the values of its sales surface are very negative), or its size leans toward a fixed value, after a series of weakened fluctuations, or it swings between boundary values with no weakening. These non-weakened fluctuations can show or not identifiable periodicities. For certain parameter values, these fluctuations are chaotic, with unpredictable results of a simulation in spite of the fact that it has been conducted from a simple and perfectly deterministic equation.

The results of these simulations are at least in part interpretable in realistic terms. For example, we notice that if the speed of adjustment is fast and/or if at the same time the demand which defines D_j is strong, the store will have a hard time reaching a stable size for certain unit cost values. It should be noted that it is possible to spatialize the model by introducing the role of other stores, based on their relative position, and the influence of the distribution of the potential clientele.

In general, the processing of dynamic models with differential equations show brutal changes of simulated behavior according to small variations of initial values and of parameters, insofar as they happen near critical thresholds. These abrupt jumps are sometimes considered an obstacle in the development of dynamic models. However, they are to a certain degree reassuring in a general way. In fact, they show that modeling integrates a very important aspect of reality, the power of small causes in producing large effects, or the importance of minor fluctuations in the interpretation of evolutions, if they happen in certain critical conditions. These minor fluctuations are sometimes difficult to insert in the framework of causality, so much so that it is tempting to say that they stem from the field of randomness; dynamic models then help to think about its part and modes of intervention, and thus constitute a decisive contribution to a major epistemological problem.



N.B. The asterisk indicates that many models of these categories have probabilistic and deterministic versions at the same time.

Figure 1.5. *A models classification based on three binary criteria*

1.5. Conclusion

Taking into account the binary oppositions described above, trying to classify models used in spatial analysis clearly shows that spatial analysis has been largely based on deterministic, aggregated, static and highly conceptual models (see Figure 1.5). This is historically true and remains so to a certain degree. Indeed, the current situation is also characterized by a large diversification: the introduction of the randomization concept from specific thematic fields such as diffusion has found several applications due to computer simulations; dynamic models help to think about new forms of the notion of cause and effect and not only about their limitations; the multiplication of uses for iconic models in reality extends the modeling field, etc.

This diversity shows that the benefits gained by each choice are always countered by drawbacks. There is no perfect modeling concept and more importantly there is not only one modeling concept. The potential choices are numerous and it is a good idea to use all possibilities offered in order to advance knowledge, without cautious or sectarian retreat on this or that space modeling method.

1.6. Bibliography

[ALL 97] ALLEN P., *Cities and Regions as Self-Organizing Systems; Models of Complexity*, Gordon and Breach Science Publishers, Amsterdam, 1997.

[BER 70] BERRY B., HORTON E., *Geographical Perspectives in Urban Systems*, Englewood Cliffs, Prentice Hall, 1970.

[BON 01] BONNEFOY J.L., BOUSQUET F., ROUCHIER J., "Modélisation d'une interaction individus, espace, sociétés par les systèmes multi-agents. Pâture en forêt virtuelle", *L'Espace géographique*, 1, p. 13-25, 2001.

[BRU 80] BRUNET R., "La composition des modèles dans l'analyse spatiale", *L'Espace géographique*, no. 4, 1980.

[BRU 87] BRUNET R., *La carte, mode d'emploi*, Fayard/Reclus, Paris/Montpellier, 1987.

[BUR 96] BURA S., GUÉRIN-PACE F., MATHIAN H., PUMAIN D., SANDERS L., "Multi-agents systems and the dynamics of a settlement system", *Geographical analysis*, 2, p. 1961-1978, 1996.

[CHA 95] CHARRE J., *Statistique et territoire*, Paris, La Documentation Française Reclus, 1995.

[CHO 67] CHORLEY R.J., HAGGETT P., *Models in Geography*, Methuen, London, 1967.

[CHR 33] CHRISTALLER W., *Die Zentralen Orte in Suddeutschland*, Iéna, 1933.

[DAU 87] DAUPHINE A., *Les modèles de simulation en géographie*, Economica, Paris, 1987.

[DEN 83] DENDRINOS D., MULLALY H., "Empirical evidence of Volterra-Lotka dynamics in US Metropolitan areas", in *Evolving Geographical Structures*, D.A. Griffith and A.C. Lea (eds.), 1983.

[DER 79] DERYCKE P., *Economie et planification urbaine*, 2 vol., Economica, Paris, 1979.

[DUR 81] DURAND-DASTÈS F., AURIAC F., "Reflections on some recent developments of system analysis in French geography", in *European Progress in Geography*, Collective, Pion, London, 1981.

[DUR 84] DURAND-DASTES F., "Systèmes et localisations: problèmes théoriques et formels", in *Géopoint 84*, Avignon University, p. 19-44, 1984.

[DUR 90] DURAND-DASTES F., "La mémoire de Gaia", in *Geopoint 90*, Avignon University, 1990.

[DUR 91] DURAND-DASTES F., "La notion de chaos en géographie", *L'Espace géographique*, 4, p. 311-314, 1991.

[DUR 95a] DURAND-DASTES F., "Les modèles en géographie", in *Encyclopédie de Géographie*, A. Bailly, R. Ferras, D. Pumain (ed.), Economica, Paris, p. 311-326, 1995.

[DUR 95b] DURAND-DASTES F., "Le Monde Indien", in *Géographie Universelle*, volume 8, Reclus/Belin, Montpellier/Paris, 1995.

[FAV 98] FAVORY F., MATHIAN H., RAYNAUD C., SANDERS L., “Sélection géographique, déterminisme et hasard”, in *Archaeomedes, Des oppida aux métropoles*, Collective, Anthropos, Paris, p. 151-248, 1998.

[GRA 96] GRATALOUP C., *Lieux d’histoire. Essai de géohistoire systématique*, Reclus/La Documentation française, Montpellier/Paris, 1996.

[HAG 65] HAGGETT P., *Locational Analysis in Modern Geography*, Arnold, London, 1965.

[HAM 67] HAMILTON I., “Models of industrial location”, in *Models in Geography*, R.J. Chorley, P. Haggett (eds.), Methuen, London, p. 361-424, 1967.

[HEN 67] HENSHALL J., “Models of agricultural activity”, in *Models in Geography*, R.J. Chorley, P. Haggett (eds.), Methuen, London, p. 443-447, 1967.

[LOW 64] LOWRY I., *A Model of Metropolis*, Rand Corporation, 1964.

[PIN 95] PINI G., “L’interaction spatiale”, in *Encyclopédie de Géographie*, A. Bailly, R. Ferras, D. Pumain (eds.), Economica, Paris, p. 539-558, 1995.

[PUM 89] PUMAIN D., SAINT-JULIEN T., SANDERS L., *Villes et auto-organisation*, Economica, Paris, 1989.

[PUM 95] PUMAIN D., “Les systèmes de villes”, in *Encyclopédie de Géographie*, A. Bailly, R. Ferras, D. Pumain (eds.), Economica, Paris, p. 645-664, 1995.

[RAV 85] RAVENSTEIN E.G., “The laws of migration”, *Journal of the Royal Statistical Society*, 1885-1889.

[REI 31] REILLY W., *The Law of Retail Gravitation*, The Knickerbroker Press, New York, 1931.

[REY 81] REYNAUD A., *Société, espace et justice: inégalités régionales et justice socio-spatiale*, Presses Universitaires de France, Paris, 1981.

[ROG 85] ROGERSON P.A., “Desequilibrium adjustment and chaotic dynamics”, *Geographical Analysis*, 17, 3, 1985.

[SAI 95] SAINT-JULIEN T., “Difffusion spatiale”, in *Encyclopédie de Géographie*, A. Bailly, R. Ferras, D. Pumain (eds.), Economica, Paris, p. 559-581, 1995.

[SAN 89] SANDERS L., *L’analyse statistique des données géographiques*, Reclus, Montpellier, 1989.

[SAN 92] SANDERS L., *Systèmes de villes et synergétique*, Anthropos, Paris, 1992.

[VON 27] VON THÜNEN J., *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, 1827.

[WEB 09] WEBER A., *Uber den Standort der Industrien*, Tübingen, 1909.

[WIL 74] WILSON A., *Entropy in Urban and Regional Modeling in Geography and Planning*, Wiley, New York, 1974.

Chapter 2

Geographical Scales and Multidimensional Statistical Methods

2.1. Introduction

Whether it be from the point of view of the biologist who works on a scale from particle to organism, or from that of social sciences and particularly the geographer analyzing events on a scale from individual to territory, notions of *scale* and *scale changes* are at the core of many discussions: we strive to understand relations or interactions between entities relative to different scales. The analysis or the modeling of these interactions has become a necessity for any discipline.

In geography, the notion of scale can be a source of ambiguity. It is always used in the sense of spatial resolution, but it can just as well refer to cartographic representation or *levels* of observation and analysis[1]. Here, we are focusing on the latter. Today, we no longer have to demonstrate that the analysis of a spatial event is directly linked to the geographical level of observation representing granularity, i.e. the *order of magnitude* chosen for the analysis. This leads to the consideration of space in relation to different orders of magnitude, each defining distinct perception levels.

Chapter written by Hélène MATHIAN and Marie PIRON.

[1] This ambiguity, inherent to geography, is not necessarily found in other disciplines.

With the development of geographic information systems (GIS), the acquisition of knowledge at different scales and the transition from one level to another are becoming technically easier. However, this creates new conceptual and methodological problems. Processing and management of data at different geographical scales will now become multidisciplinary, involving computer scientists, statisticians, geographers and other scientists working with space. The focus has been mainly on modeling of hierarchized spatial data in terms of databases [RAY 96] and it would seem that, from a processing standpoint, the integration of formalizations and statistical methods adapted for this type of data in the different tools is still weak [OPE 96, REY 98]. Statistically, it is true that geographical data could not easily adapt to the constraints of assumptions on which traditional statistical models rely (independence between observations, identically distributed, etc.). However, for a certain number of years now, specific methods for the processing of structured data have been developed; the localized data are remarkably representative of such structured data. Geography should therefore be a major application field. More generally, multilevel methods have emerged in social sciences and contribute to the connection of the different levels and both observation levels and organization levels.

We wish to position ourselves at the crossroads of these thoughts and present different processing methods integrating the *multilevel* aspect of space, without losing sight of the geographic questioning. In section 2.2, we first propose the distinction between two methods that seem to emerge from different issues: one prioritizing the connection between a set of nested levels and thus addressing the question of the change of level and the other prioritizing the integration of global levels to a lower level within the frame of contextual analysis. Starting with these methods, we are setting out to provide statistical formalization useful for the creation of multilevel information as well as for the application of adapted statistical methods. The applications then presented explain analysis techniques emerging from descriptive and explanatory methods. The former, which are older but less known, apply very well to spatially structured data under partition graphs or, by extension, proximity graphs. These methods are generic enough to deal with issues relative to both approaches described earlier: the connection between levels (see section 2.3) and the consideration of context effects (see section 2.4). The latter multilevel models presented in section 2.5 are specifically dedicated to the integration of contextual effects in more traditional statistical methods.

In works on multilevel models, the terms “multilevel” and “contextual” are deliberately reserved for explanatory methods, but we will use them here in a more general way to qualify multivariate methods integrating information of different levels and in particular, information relative to contexts.

2.2. Scaling issues

Multidimensional statistical methods, which are appropriate for the analysis of a set of structured information over several geographic levels, require a certain formalism that we address in this section before presenting the methods themselves. We start with a formalism from geographical questioning, to which is then associated a statistical formalism of different relations that we need to consider.

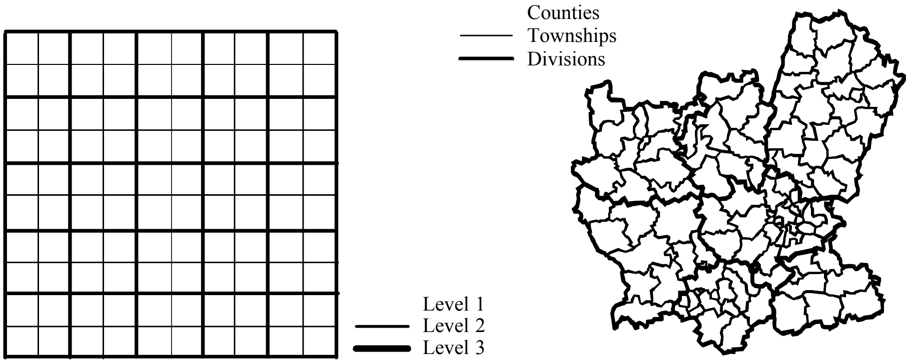
2.2.1. *The consideration of different geographical levels: two possible approaches*

"*The notion of scale is connected to the notion of measure of space or time. Going from one scale to another means changing units of measure*" [PAV 94]. Thus defined, this notion refers to an infinite number of map zonings in regular and homogenous grids, constituting the "unit of measure" for the observation of space (see Figure 2.1a). This notion of scale concerns all those perceive space as a continuum such as physicists, ecologists, etc. In social sciences and especially in geography, the notion of scale mostly refers to a set of predefined levels often forming irregular and non-homogenous partitions of space and leaving very little room for the choice of order of magnitude of the grids (see Figure 2.1b). Such partitions can have coherent functioning logic for the event studied, or can be only mandatory passages for the collection of certain information.

At all events, grids often show significant size and form heterogenity. These map zonings are called into question because they introduce strong biases on observed spatial relations and organizations as well as on their interpretation [FOT 91, GRA 01, OPE 84]. This well known problem is called a *modifiable areal unit problem* (MAUP) and is characterized by two effects. The first, called "scale effect", is linked to the bias introduced by the choice of a level for the observation of a distribution. The second, called "zoning effect", is linked to the bias introduced by the choice of a specific zoning for a given spatial level[2].

---

[2] Although this issue seems of particular importance to us, we do not explicitly address it here.



a) Continuum of levels b) Levels defined *a priori*

Figure 2.1. *Notion of spatial scale*

Until the 1980s, statisticians used to try to control these effects, but today it is accepted that it is much more useful to try to integrate them. We try to surpass the scale defined *a priori* by territorial contexts and to integrate the continuous character of space. This requires methods that will take into consideration the multilevel dimension of geographical data. Two types of approaches can be distinguished.

Approach concerning level changes

We are focusing here on a set of nested levels constituting the scale and the emphasis is put on their *articulation*. Covering different levels effectively produces a change in the perception of the studied event; for the higher levels, details disappear to make way for global forms and larger expanses, whereas the lower levels reveal local spatial differentiations [DUR 99]. Zoning acts as a filter affecting the significance of the event's measure [RAY 96]. The change of observation levels can have heuristic dimensions; it can: define the different organization levels of an event, update new spatial structures and new spatial objects, and focus on the relevance of a geographical zoning, given or arbitrary, by taking the internal variability of spatial units into account. Going from one level to the other implements aggregation procedures which transform information, as this information is not analyzed by its originating primary unit anymore, but by an aggregated unit. We then expose ourselves to the risks of an ecological fallacy[3]. This approach attempts to explain the connection between the different nested

---

[3] The ecological fallacy is an interpretation error consisting of inferring results obtained at a certain level to a lower one.

partitions in the territory and focuses on the comparison of spatial organizations linked to these different levels.

Approach concerning contextual effects

We concentrate here on the analysis of a specific level, generally the lowest one. The units of this level are analyzed in relation to larger spatial entities conditioning them and defining *contexts*. The notion of context is completely accepted when the object studied is the person and we are trying to integrate his social or spatial environment in the analysis. When the object of study is a spatial entity, we make the assumption that the places must be considered in relation to their environment. Indeed, the study of an event in a given geographical level can return a certain number of mechanisms relative, on the one hand to processes operating at the studied level, and on the other hand to processes that bring interactions between places into play, based on ranges that differ and that will refer to different levels [SAN 99]. The purpose is to account for the variability of observed units considering the variability of their environments. The consideration of the context helps to compensate for the individualistic fallacy[4] which is well known by those working at individual level.

2.2.2. *Formalization of relations between two levels*

In these two approaches, the focus is on neighborhood and nested relations between entities of different geographical levels. These relations can be represented with a common mathematical formalism in order to homogenize the statistical analysis process.

We note by I, $I = \{i, i = 1 \ldots n\}$ the set of n statistical units of the lowest level, considered as primary entities, aggregated or not, and which cannot be broken down for the study. We note by P, $P = \{p, p = 1 \ldots m\}$ a set made up of m entities, each entity being a class of I entities. P entities are larger area entities than those of I, thus P constitutes a higher level.

2.2.2.1. *Nested relations and partition graph*

Talking about nested relation is the same as saying that I and P form two partitions of the studied space, such that each entity i of I belongs to only one class p of P.

---

[4] The individualistic fallacy consists of describing and analyzing people in relation to their own characteristics, independently of their environment and of the different contexts that will influence their behavior.

The most obvious way to explain this relation is to describe the set of entities of *I* with one structure variable illustrating the data organization. The goal is the entity identification of level *P*, corresponding to the partition variable or still a qualitative variable to *m* terms associated with each *m* classes of *P*. It is usually transformed for procedures of aggregation and for statistical processes in two matrices:

– an affiliation matrix, K^P. It is a complete disjunctive table such that the general term K_{ip}^P has the values 1 or 0 depending on whether unit *i* belongs to class *p* of *P* with $i = 1, \ldots n$ and $p = 1, \ldots m$ or not:

$$K_{ip}^P = \begin{cases} 1 \text{ if } i \in p \\ 0 \text{ otherwise} \end{cases} \qquad [2.1]$$

where $K_{i.}^P = \sum_p K_{ip}^P = 1$ and $K_{.p}^P = \sum_i K_{ip}^P = n_{.p}$ is size of class *p* of *P*.

This type of formalization will control the connection between levels *I* and *P*; *the identifications of P's* class remains. It favors level *P* by keeping the composition of its units;

– a symmetric matrix G^P, of general term $G_{ii'}^P$, with $i,i' = 1, \ldots n$, resulting in the relation of equivalence that partition *P* defines on *I*. With the *(i,i')* pair, we associate the value 1, if *i* and *i'* belong to the same class *p* of higher level and value 0 otherwise:

$$G_{ii'}^P = \begin{cases} 1 \quad \text{if } i \text{ and } i' \text{ belong to the same class } p \text{ of } P \\ 0 \text{ otherwise} \end{cases} \qquad [2.2]$$

where $G_{i.}^P = \sum_{i'} G_{ii'}^P = n_p$ is the size of class *p*, but also the number of elements of *I* in relation to *i*. G^P is the matrix associated with the graph of partition *P*. It is important to note here that we lose the identification of higher level classes. This formalization is used when we wish to favor level *I* and the structure that the higher level defines on this set, without having to identify higher level classes. That is the case with local analyses (see section 2.4).

In practice, we often have to make several partitions that do not necessarily arrange themselves according to a nested relation, such as, for example, the partition of family allowance funds or chambers of commerce sectors (see Figure 2.2), where partitions are presumed incompatible. To analyze them, it is necessary to come down to a common level, either by aggregation (department level, for example), or by disaggregation (municipality level).

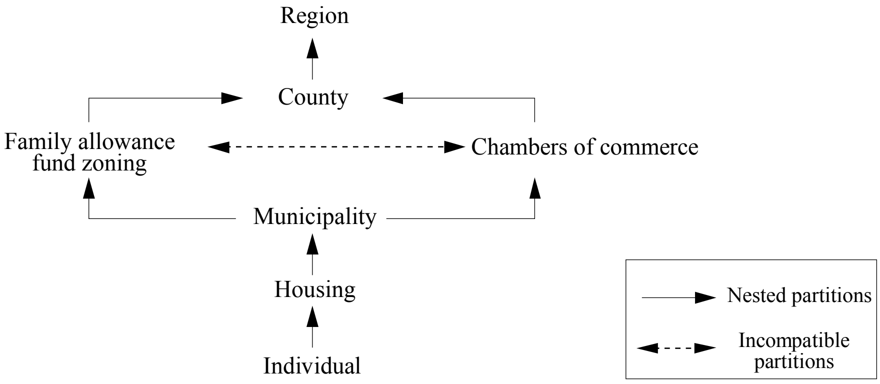


Figure 2.2. *Nested and incompatible map zonings*

2.2.2.2. *Neighborhood relations and proximity graphs*

This is the case where a defined relation over group *I* builds higher level entities by aggregation, without them becoming a partition of *I*. In a broad sense, this type of relation is called *neighborhood relation*. Let *R* be this relation; the neighborhood of an entity *i* of *I* is the group of entities that are related to it, so $V(i) = \{i' \in I \,/\, i\, R\, i'\}$. This relation can be defined from a measure of proximity or distance[5] δ in space, or of a functional relation based on proximity relations in space. Figure 2.3 illustrates a hierarchy where municipalities are grouped according to inter-municipality cooperations to which they belong to. These unions are associated with functional relations between the towns.

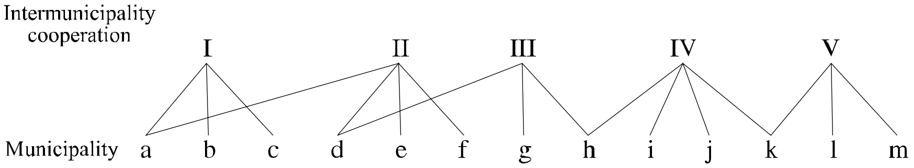


Figure 2.3. *Hierarchy associated with a neighborhood relation*

---

[5] By analogy with the notion of distance, we will name it *d*, although often it is not really a distance in the mathematical sense.

For relations completely defined by proximity in space, performances of GIS today enable a large diversity of relations and the implementation of separation measures associated with them. We can put them in three categories:

– adjacency relations: $i \, R \, i'$ if i and i' are adjacent;

– measures using geometric distances: $i \, R \, i'$, if $\delta \, (i,i') < r$, where r is a threshold;

– those dealing with geometric calculations like the intersection with another group Z of geographical objects: $i \, R \, i'$, if $\exists \, z \in Z \, / \, i \cap z \neq \varnothing$ and $i' \cap z \neq \varnothing$. ($Z$ can represent the sections of a transport network, for example, whereas I designates a group of municipalities).

The set of neighborhoods forms a new level V, which is made up of m entities. They do not necessarily constitute a partition of space. Based on the chosen relation R, the number of classes V can be very high, even equal to or higher than the number of units i of I. The hierarchization of levels is ensured by the affiliation relation of the entities of level I to one or more classes of level V. As for the case of partitions, which is particular, two matrix formalizations can be useful:

– an affiliation matrix K^V which is a logical table describing the affiliation of entities I to entities V. For each distinct "neighborhood" entity, an indicating variable will be associated with a value of 1 for its units and 0 otherwise. We designate K_{iv}^V as the general term with $i = 1, \ldots n$ and $v = 1, \ldots m$

$$K_{iv}^V = \begin{cases} 1 \text{ if } i \in v \\ 0 \text{ otherwise} \end{cases} \qquad [2.3]$$

where $K_{i.}^V = \sum_v K_{iv}^V = n_{i.}$ is the number of classes or neighborhoods V to which i belongs and $K_{.v}^V = \sum_i K_{iv}^V = n_{.v}$ the neighborhood asset v of V;

– matrix G^V expressing the neighborhood or proximity relation between two units of I. The general term $G_{ii'}^V$ with $i,i' = 1, \ldots n$, is valued at 1 if i' is in the neighborhood of i and 0 otherwise:

$$G_{ii'}^V = \begin{cases} 1 \text{ if } i \text{ and } i' \text{ are neighbors} \\ 0 \text{ otherwise} \end{cases} \qquad [2.4]$$

where $G_{i.}^V = \sum_{i'} G_{ii'}^V = n_i$ is the neighborhood asset i, but also the number of units I in relation to i. It is the matrix associated with the proximity graph defined by the neighborhood relation of V. Matrix G^V is usually asymmetric and because of this, the weight of two units in relation (row and column spreads) are different. We can qualify this formalization and introduce different proximity intensities with a weighting system [CLI 73]. In this way, if *i'* is a neighbor of *i*, $G^V_{ii'}$ represents the weight of *i'* in the neighborhood of *i*.

Most often, we would wish that neighborhoods have the same weight of *1*. We then divide rows of G^V by their respective weights [BEN 00]. By noting by D^V the diagonal matrix of dimension (*n,n*) such that $D^V_{ii} = 1/n_i$, we are working on matrix $D^V G^V$ such that $(D^V \mathrm{G}^V)_{\mathrm{ii'}} = G^V_{ii'}/n_i$.

By extension, entity neighborhoods of *I* can be defined on any group of entities *Z*, which are adapted to the question asked, like $V(i) = \{z \in Z \,/\, i\, R\, z\}$. In [SAN 00], for example, the neighborhood of a city is made up of a group of urban units located at less than 25 miles. In this way, we can classify the neighborhoods of a group of entities as hybrid contexts, which are made up of other types of entities than those from the group studied. In this case, only the affiliation table or the complete disjunctive table K^V can be produced.

2.2.3. *Processing of multilevel information*

The formalization of the elements of a multilevel structure requires the identification of not only the entities relative to the different levels and their relations (nested, neighborhood), but also of the nature of the attributes describing them. This is vital because the multilevel information is made up of a group of measures which most often result from statistical aggregation procedures. When they are implicit, they are often a source of ambiguity in the construction of data, the use of the method and the interpretation of results.

2.2.3.1. *Multilevel structure and attributes*

Let us take the example in Figure 2.4 [DAU 94] which illustrates a hierarchical organization of nested levels in which D'Aubigny distinguishes:

– the primary level (level $h = 0$), whose entities are called *atoms*, which is the lowest level of the hierarchy for which the information is available;

– the collection level (level $h = -1$), which is the level where the measures were actually taken.

When we have access to the collection level, these two levels can be confused. In geography, when the studied units are spatial entities, these levels are most often distinct. The measures associated with spatial entities (primary level) are done on lower level entities (collection level) that are not introduced in the multilevel information system. For example, in the analysis of the municipalities dynamics, the atom is the municipality. However, if the analysis is done on the populations of the towns, the residents constitute, in an implicit way, a non-measurable lower level. On the other hand, in a system of inquiries with multiple levels of observation (person, household, housing), all levels are measurable and we can have several collection levels. The lowest level of collection is the primary level.

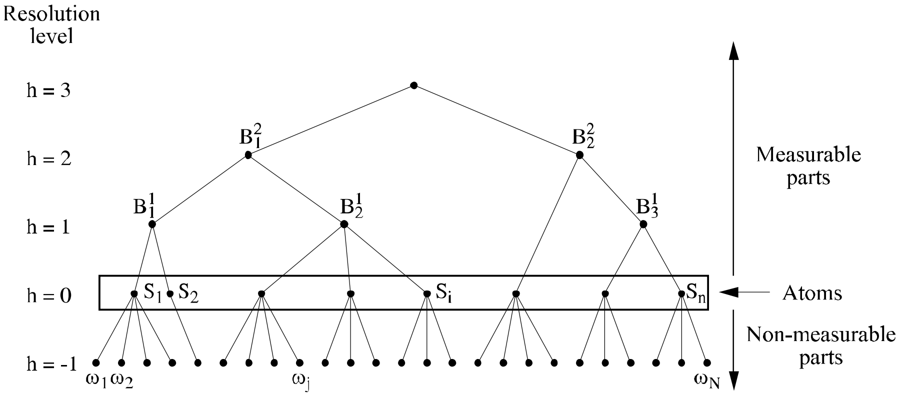


Figure 2.4. *Measurability and hierarchy of levels (source: [DAU 94])*

The *aggregation procedure* is not an insignificant operation: it can transform not only the nature of the attribute itself – for example, the aggregation of a qualitative variable results, at the higher level, in as many quantitative variables as there are methods – but also the semantic – average, weighted average, predominant rule for a quantitative variable.

Each aggregation procedure requires determining to which study population and to which level of the hierarchy the attribute refers. Based on the aggregation procedure chosen, we can have a reaggregation. That would mean the measure created at the higher level does not describe the population associated with the next lower level, but a population of another level down, most often the collection level. It is, for example, the difference between an average and a weighted average. The introduction of a weighting system for the statistical units of each level of the scale enables us to keep the reference to the population collected. This methodology is generalized on several collection levels in [DAU 94, PIR 92].

2.2.3.2. *Multidimensional statistical methods*

Higher level entities of a hierarchy of spatial levels constitute primary entity aggregates that are not homogenous according to the studied variables. These occur in particular when entities correspond to an arbitrary geographical partitioning. This internal heterogenity affected by the aggregation procedures can lead to false interpretations of which we are still trying to control the effects [CHA 95, OPE 84, PIR 93]. This problem has caused several methodological developments that we can group in two basic directions. Some are trying to resolve the problems specifically linked to effects of scale and mesh (MAUP) and quantify them with the purpose of minimizing the risks of false interpretation. They are mainly focused on a methodological research process and propose "adjustments" to methods, which are based mainly on inferential statistical analyses (for example, see [WRI 96]). GISs have in the last 10 years theoretically enhanced these developments and helped to advance experimentation [REY 98]. Parallel to this direction, a second approach takes multilevel structure of localized data into account and considers that the set of levels may have an effect on the analyzed level. The set of levels is considered again and integrated in the analyses in order to constitute as many explanatory factors [MAR 99]. We focus on this type of approach and will present three thematical applications illustrating this.

In geography it started with applications using traditional and simple methods. For example, the analysis of the variance constitutes the simplest method, and also the most used, to identify the "context effects" and to measure the internal heterogenity of a set of aggregates. Homogenity (or heterogenity) indicators are developed from different relations between global variance, interclass variance or intraclass variance [HAG 65, PUM 97, SAN 89]. Similarly, in the case of contexts, experiments have been attempted in order to process multilevel information with traditional methods (for example, see [DUP 94, ROZ 89]), without considering the relations between levels or between entities of a same level, the relation between levels being done *a posteriori*.

A multilevel structure of information is by nature multidimensional, as much from the number of variables to take into account as from the levels to consider. Among the multidimensional statistical methods adapted to structured data over several levels [LEB 06], we have focused on *local analyses* that come from a more descriptive approach and *multilevel models* that emerge from an explanatory approach. The three following sections are applications of these different methods.

In section 2.3, we favor the analysis of a change of scale in an exploratory way. Since no level is fixed *a priori*, we are trying to bring out the general trends for each aggregation level and to analyze structure modifications in relation to the lower level. Section 2.4 illustrates a contextual descriptive analysis where the different higher levels are integrated in the analysis of the lowest level. We are trying to eliminate the general trends to only take into account the effects linked to contexts. Finally, in section 2.5 we show how an explanatory method, in this case a multilevel model, developed for the analysis of effects of context in socio-demographic studies can be interesting for the process of geographical issues and complement the analyses presented in section 2.4.

2.3. Change of levels, change of structures

With the idea of change of level, it is interesting not only to bring out structures specific to each level of the scale, but also to extract the structures of the change itself when we go from one level to another. These structures characterize the way spatial units are organized and also make it possible to control the aggregation procedure. The goal is then to address such questions as: what is the influence of the chosen geographical level over the perception of a given phenomenon? How is this perception modified over all the levels? Is there an adequate analysis level for the observation of an event? At which level does a variable lose or gain a power of discrimination? Will we find the same organization elements at the different levels? What is the nature of the internal heterogenity of the spatial units? Throughout this series of questions, the goal is to understand if we are in the presence of a phenomenon measured at different levels or if at each level there are different phenomena.

Let us now consider the transitions between the different levels of the scale in order to facilitate an integrated reading of the different factors at different levels. In order to do this, we have chosen to present in more detail an exploratory analysis method relative to changes of geographical scales [PIR 92]. This method includes different methodological aspects which are common to several other methods (aggregation formalism, interclass and intraclass analyses, homogenity indicators) where the attributes are qualitative. To illustrate our point, we are using the study of the urban extension of Ouagadougou (Burkina Faso) done in 1987. We are attempting to define the organization of the outlying space of the city and to update the social and spatial patterns from the characteristics of the population and urban politics. This encompasses the period for systematic allotments of the city’s outlying districts operations. In order to complete this study, a survey at several observation levels has been done over roughly a thousand households residing in and around the parcels that they occupy [LEBR 92].

2.3.1. *Scale and variability*

The considered scale is made up of four observation and analysis nested levels. Two groups of attributes are observed at the lowest levels:

– household (*M*) is the primary level that cannot be broken down. It is the collection level made up of 958 households. It is the first level of attribute definition giving socio-demographic characteristics of the households and their integration to the city;

– parcel (*P*) corresponds to the first aggregation level. Households surveyed are spread over 657 parcels. This level provides the second level of attribute definition and we measure the characteristics of the parcel, the access to the land and to housing;

– census zone (*Z*) is an arbitrary spatial division, of 188 zones, put in place for the census survey of the population in 1985;

– sector (*S*) is an administrative division which groups census zones. It was put in place to replace the old traditional district sectioning. The periphery is divided in 16 sectors.

The chosen study population is the group of households. We have a hierarchical structure of the information available and the scale system used is complete since the primary level, the household, corresponds to the lowest level of data collection.

The inclusion relation between levels brings on the study population a series of nested partitions, which are relative to the different arbitrary map sections (census zones) or measured (residential groups), constituting zones or groups of households somewhat homogenous. In this way, the global variability of households is broken down in variability internal to classes and variability between the classes based on the different levels (Huygens' principle: see Box 2.1).

2.3.2. *Exploratory analysis of the scale system*

The goal of the method is to achieve two types of complementary analyses:

– aggregation levels analysis, which accounts for the average profiles of aggregated entities and which explains the interclass inertia;

– on the other hand, the analysis of the transition between successive levels, which accounts for the heterogenity of the profiles of lower level entities inside their affiliation class (i.e. higher level entities) corresponding to the intraclass inertia.

Box 2.1. *Breakdown of inertia based on the principle of Huygens*

We note by M (respectively P and S) the set of households i (respectively for parcels p and sectors s). By considering the set of households as a scatter plot on which the partition in Parcels acts, the inertia of the plot called *In(M)*, is broken down in interclass inertia, *In.Inter(P)* (variance between parcels) and intraclass inertia, *In.Intra(P)* (variance of households inside their affiliation parcel). For the double Parcel-Sector partition over the group of households (see Figure 2.5), the breakdown of inertia becomes, under the principle of Huygens:

$$In(M) = In.Inter(S) + In.Intra(PS) + In.Intra(P)^{6} \qquad [2.5]$$

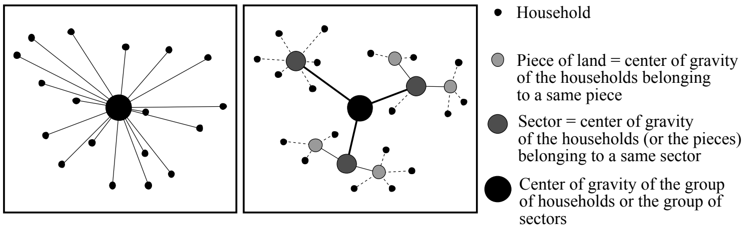
If g represents the scatter plot's center of gravity, g_p and g_s the centers of gravity of the sub-plot of households respectively from parcel p and sector s, and m_i, m_p and m_s, the respective weights of households i and of parcel p and s, each of the terms are then $i \in M$, $p \in P$, $s \in S$ for everything

– $In(M) = \sum_{i} m_i d(i,g)^2$;

– $In.Inter(S) = \sum_{s} m_s d(g_s, g)^2$ is the intersector inertia and measures the degree of heterogenity between sectors;

- $In.Intra(P) = \sum_{p} \sum_{i \in p} m_i d(i, g_p)^2$ is the intrapiece of land inertia and measures household variability inside the parcel;

- $In.Intra(PS) = \sum_{s} \sum_{p \in s} m_p d(g_p, g_s)^2$ is the intrasector inertia over the group of parcels (and corresponds to the intraclass inertia between two levels of aggregation). This term measures the variability between parcels (i.e. household classes that is the study population chosen) in relation to their affiliation sector. Since the basic elements, parcels, are already aggregates, it illustrates the transformation of the information in the transition between two aggregation levels.



Inertia of the group of households = *In.Inter(S)* + *In.Intra(PS)* + *In.Intra(P)*

Figure 2.5. *Breakdown of inertia based on the double parcel-sector partition over the group of households*

We can easily generalize this breakdown to a series of nested partitions $\{P_1, ..., P_m, ..., P_n\}$ over the finite group of households M with:

$$In(M) = \sum_{n} In.Intra(P_{n-1}P_n) \qquad [2.6]$$

where $In.Intra(P_{n-1}P_n) = In.Inter(P_{n-1}) - In.Inter(P_n)$ [7]

---

[6] We note that $In.Inter(S) + In.Intra(PS) = In.Inter(P)$
and $In.Intra(PS) + In.Intra(P) = In.Intra(S)$.

This series of analyses of aggregation levels and of transitions between these levels (see Figure 2.6) provides a global vision of the spatial organization of a phenomenon and of the transformation of structures as we change geographical levels. When we have a primary level of collection (household level), the analysis of this level provides the organization of raw data (see Figure 2.6a). Since the attributes are qualitative, we perform a series of correspondence analyses[8] on each of the tables associated with the terms of inertia breakdown presented in Box 2.1 and in [BENZ 73].

2.3.2.1. *Analysis of aggregated levels or interclass analysis*

At each level of the scale, we perform an analysis of the interclass inertia. Each basic statistical unit, the household, is studied relatively to the center of gravity of its affiliation class, parcel (see Figure 2.6b) or sector (see Figure 2.6d). In this way we analyze the average behavior of each parcel or sector. For each aggregation level, we highlight the associations or differentiations between their own spatial units. Each level has an image of the social structure of the urban extension associated with it.

For each aggregation level, we normally establish the table that will intersect the partition variable, which corresponds to the set of entities of the considered level, with the group of attributes J describing these entities. In the case of qualitative attributes, this table becomes a contingency table. For the sector level, for example, it is called T^S (see Box 2.2) and is obtained either:

– by aggregation from a complete disjunctive table K_{MJ} of size (M,J)[9] relative to the household level and from the affiliation table K^S_{MS} (see section 2.2.2): $T^S_{SJ} = K'^S_{MS} K_{MJ}$ and T^S_{SJ} is of size (S,J); or

– by reaggregation from a contingency table T^P_{PJ} corresponding to a lower aggregated level than the one considered, $T^S_{SJ}=K'^S_{PS}T^P_{PJ}$. The study population remains the one defined by the primary non-aggregated collection level, i.e. the household[10].

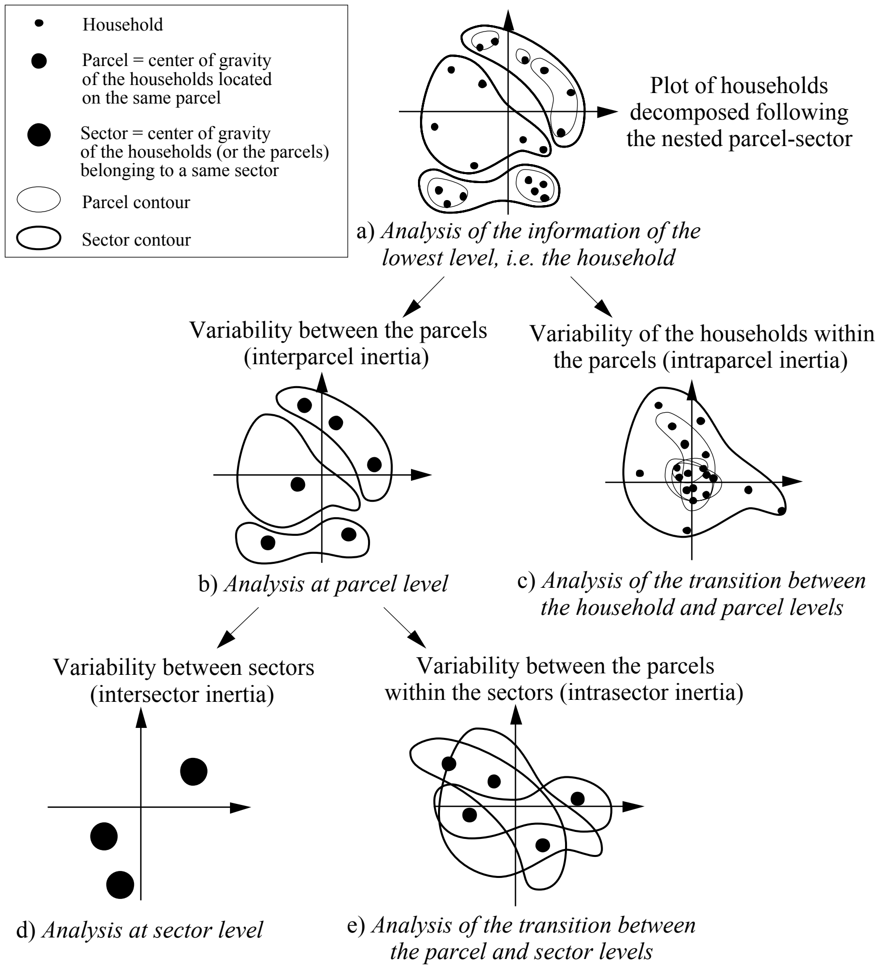
---

[7] This term can be identified with the level index of an ascending hierarchical classification, whose goal is to organize data by creating a relation defined by an interclass inertia maximization aggregation criterion. This comment contributes to the characterization of a nested levels system as a classification process defined *a priori* and describing data organization determined by the existence of a hierarchical relation.

[8] For quantitative attributes (ratios, measures, etc.), we can perform a principal component analysis (see section 2.4).

[9] The matrix indexes, $M = card(M)$, $P = card(P)$, $S = card(S)$ *and* $J = card(J)$ indicate matrix sizes.

[10] We understand here the importance of constantly questioning on which statistical population and at which level the process must be performed.



a) *Analysis of the information of the lowest level, i.e. the household*

b) *Analysis at parcel level*

c) *Analysis of the transition between the household and parcel levels*

d) *Analysis at sector level*

e) *Analysis of the transition between the parcel and sector levels*

Figure 2.6. *Series of inter and intraclass analyses on all households based on household-parcel-sector nested*

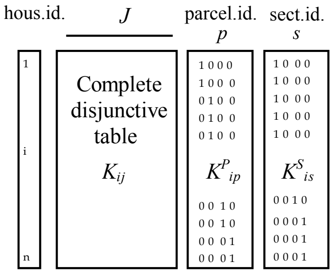
Note that the analysis of table T^S_{SJ}, corresponding to the sector level, for example, is equivalent to the analysis of table $T^S_{MJ}=D^S_{MM}G^S_{MM}K_{MJ}$ (where D^S and G^S are respectively the matrix of the inverse of weights and the matrix associated with the partition graph *S* over *M* (see section 2.2.2)). It has the same size (*M,J*) and the same margins than the complete disjunctive table K_{MJ} and each of its rows corresponds to the unit average of one class. This analysis is also equivalent to the one in table $T^S_{PJ}=D^S_{PP}G^S_{PP}T^P_{PJ}$ for partition *S* over *P*. In the same way, the analysis of the correspondences of tables T^P_{PJ} and T^P_{MJ} is equal. These types of tables are useful for the transition analysis between two levels.

2.3.2.2. *Transition analysis between two levels or intraclass analysis*

The transition analysis between two levels corresponds to the analysis of the intraclass inertia where each statistical unit of the lower level (primary or aggregated) is represented by the difference with the average of its class. We study the dispersion of statistical units inside their spatial affiliation unit.

The transition analysis between household and parcel levels (see Figure 2.6b) (respectively between parcel and sector levels) must show the differentiation of households inside parcels (respectively of parcels inside of sectors). This means representing a statistical unit (primary for the household or spatial if taken from an aggregated level) not by its deviation with respect to the general average of the studied population as in a traditional factor analysis, but by its deviation with respect to the population average belonging to the same unit. From the table showing the lower level of analysis (the complete disjunctive table K_{MJ} if it is the primary level M or the aggregated data table T^P_{MJ}, which is equivalent to the geographical information table T^P_{PJ} and relative to level P), we cut out the table corresponding to the higher aggregation level respectively T^P_{MJ} for parcels or tables T^S_{MJ} (equivalent to T^S_{PJ}) for sector[11]. Such a table, shown as Δ, results in the intraclass inertia and measures the share that each unit can restore from the commonly defined structure.

Box 2.2. *Construction of information tables relative to the different aggregation levels parcel and sector – qualitative variable case*



For all $i \in M, p \in P, s \in S,$ we have:

– at parcel level:

$$T^P_{pj} = \sum_{i \in I} K_{ij} K^P_{ip} = \sum_{i \in p} K_{ij}$$

– at sector level:

$$T^S_{sj} = \sum_{i \in s} K_{ij} = \sum_{i \in I} K_{ij} K^S_{is} = \sum_{p \in s} \sum_{i \in p} K_{ij} = \sum_{p \in s} T^P_{pj}$$

---

[11] We can add the product of margins $K_M K_J / k$ (where k is the population size) corresponding to the assumption of classic independence, in order to implement the usual program of correspondence analysis (we verify that all tables K, T and Δ for the same levels have the same size and margins).

The analysis of the transition between the primary collection level *M (households)* and the sector level *S* is an analysis of multiple conditional correspondences introduced by Escofier [ESC 87]. The correspondence analysis appears in the following equation:

$$\Delta_{MJ}^{MS} = K_{MJ} - T_{MJ}^{S} + K_M K_J / k \qquad [2.7]$$

where *k* is the population size.

In order to study the transition between two aggregated levels, parcel *P* and sector *S* (see Figure 2.6e) representing the most common case in geography, we generalize the analysis of multiple conditional correspondences by making a correspondence analysis in the following equation:

$$\Delta_{PJ}^{PS} = T_{PJ}^{P} - T_{PJ}^{S} + T_P^P T_J^P / t \qquad [2.8]$$

This type of analysis explains the structures used by the lower level and that change during the change of geographical scales.

2.3.3. *Application of outlying Ouagadougou space to the social and spatial organization*

In order to highlight the main sociological and landed elements of the social and spatial organization of the Ouagadougou periphery, we apply the analysis of the scale system. We will present the main trends provided by the principal factorial plans of each analysis (see [PIR 92] for more detail).

The series of scales system analyses (see Figure 2.7) shows a progressive modification of dominating structures. Those obtained at Household and Sector levels are very different. Several movements are coming up in reading the scale system organization.

At household level (see Figure 2.7a), three household profiles, corresponding to the life cycle emphasized by the main plan, structure Ouagadougou's social organization:

– "students" (young, single, schoolchildren or students) do not own properties and recently arrived at Ouagadougou and on their parcel;

– "modern" households, monogamous, homeowners or tenants of recently attributed parcels, where the head of the household is middle aged and mobile;

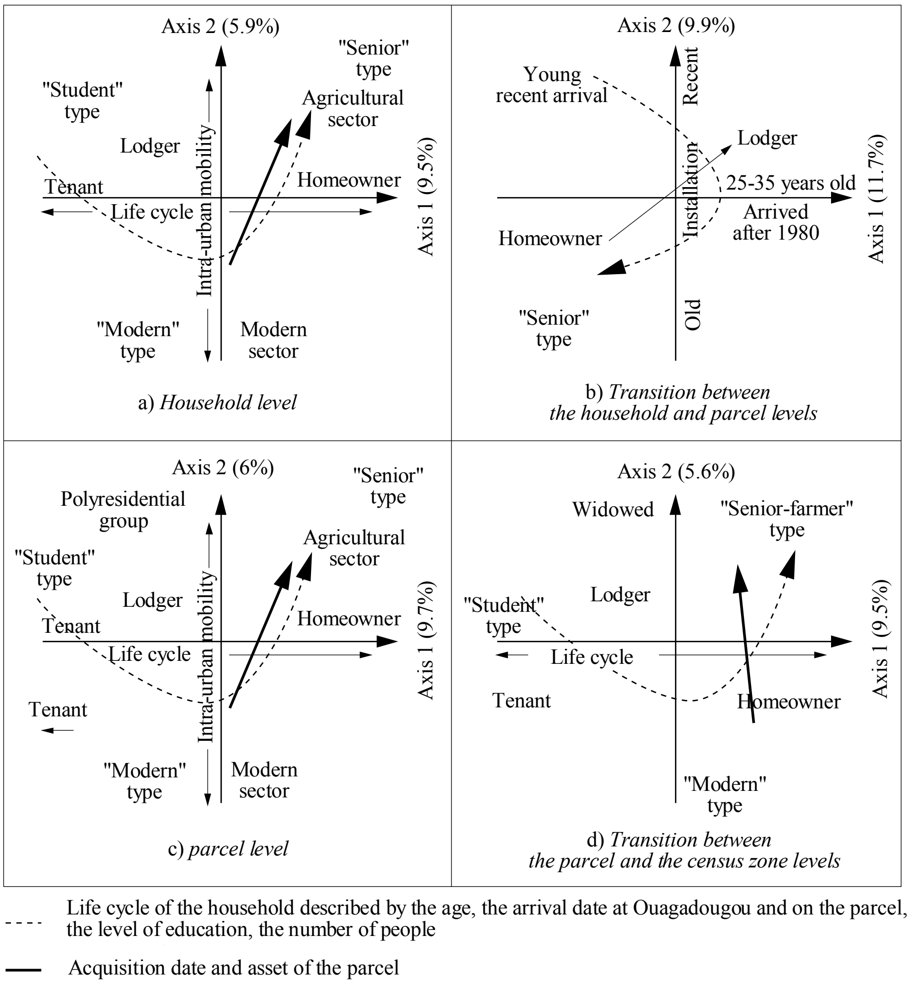
– “senior” households, highly represented in the agricultural sector and among natives to Ouagadougou, have been in this city for a very long time and are not very mobile; the head of the household is old, polygamous or widowed, uneducated and has acquired his parcel in a traditional way.

The structure obtained at household level is held at parcel level (see Figure 2.7c): this level is made up of very homogenous pieces of land in terms of household composition (which is obvious as the majority of pieces of land have only one household). We note that the polyresidential groups (household groups with the same profile or different profiles and living in the same parcel) are mostly made up of the young and senior households.

The analysis of the transition between household and parcel levels (see Figure 2.7b) has in fact a very different configuration than those analyses performed at each of these levels (and a very weak intraclass inertia). It emphasizes households with different profiles and living together: young households recently arrived at Ouagadougou, “senior” households, long time city homeowners. It logically excludes monoresident households and polyresidential groups whose households have identical profiles (for example, student households). It emphasizes the residential cohabitation “homeowner-lodger”.

The analysis of transition between parcel and zone levels (see Figure 2.7d) globally accounts for the same main structures than that at parcel level (therefore household level). It emphasizes parcel and household profiles which co-reside in census zones and that become elements of heterogenity and dispersion at this level. For example, the “student-pupil” type is distributed over all census zones and therefore is no longer a visible and decisive element of this level (see Figure 2.7e); it distinguishes itself from the profile of the “student-tenant” continuing which is graduate studies and located in the university’s sector in the East.

At census zone level (see Figure 2.7e), the analysis mainly highlights a relation between urban fabric restructuring (nature of allotment operations) and residential structures (acquisition modes and parcel installation date) with certain socio-economic structures defining marginal household profiles such as “senior-Ouagadougou resident” or “student-tenant”. This last profile is made up of well located homogenous polyresidential groups (and thus that could not come out from the analysis of parcel internal variability). We also note the more important presence of homeowners in zones undergoing complete restructuring. Globally, the main trends in the organization of space inherent to the census zone scale are: 1) south of Ouagadougou, a majority of households recently arrived in the city and on their parcel that is allotted or not allotted; 2) on the land north of the city undergoing restructuring, a majority of previously installed households in Ouagadougou and on their parcel acquired traditionally; 3) dispersed over the whole periphery, modern and recent attribution parcels where we do not find a dominating household type.



a) *Household level*

b) *Transition between the household and parcel levels*

c) *parcel level*

d) *Transition between the parcel and the census zone levels*

Figure 2.7a-d. *Scale system analyses*

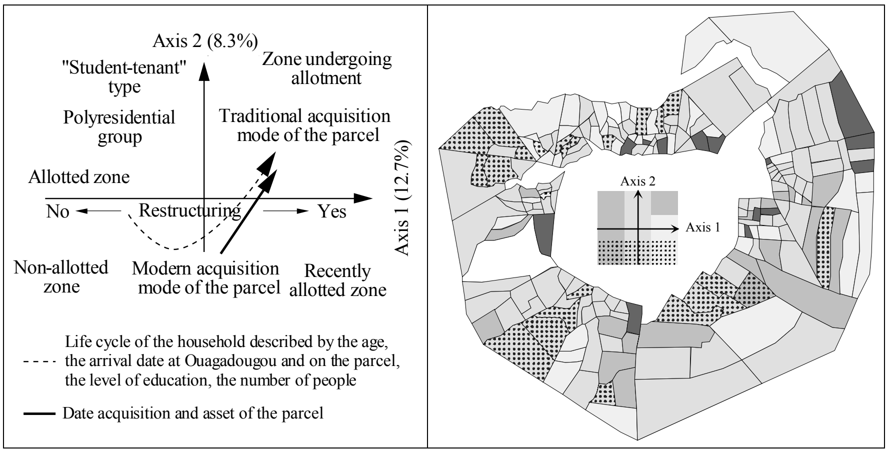


Figure 2.7e. *Scale system analyses – census zone level*

The analysis of the transition between the census zone and sector levels (see Figure 2.7f) highlights the same household profiles as the analysis of the census zone, "senior" profiles (farmer and/or Ouagadougou resident) and those residing on modern and recent allotment parcels. Spatial differentiation of socio-economic structures is no longer eliminated at sector level. The corresponding map makes it possible to represent the degree of heterogenity of the sectors and to differentiate the generally homogenous sectors with regard to census zone profiles.

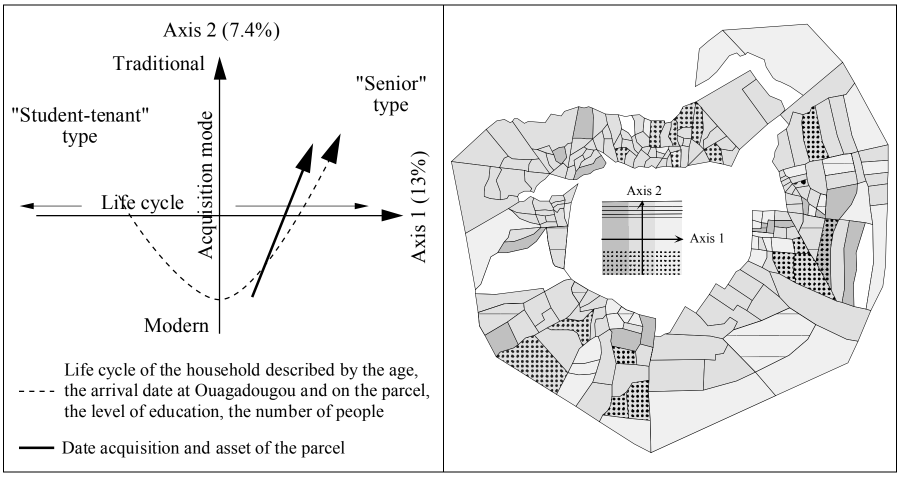


Figure 2.7f. *Scale system analyses – transition between zone and sector levels*

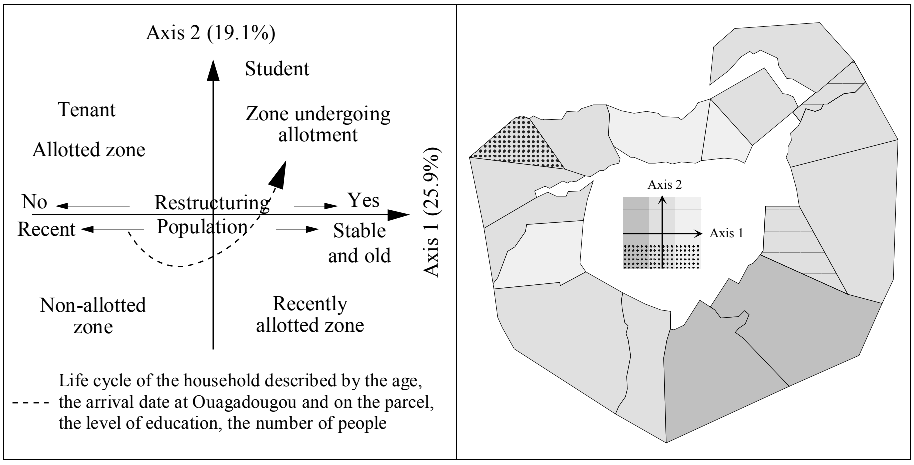


Figure 2.7g. *Scale system analyses – sector level*

At sector level (see Figure 2.7g), the analysis is not structured like that at the lower level, the census zone level (see Figure 2.7e). There is no relation between planning operations and a social organization on the periphery of Ouagadougou: in other words, when the study was done (1987), the planning operations had not generated a social segregation process. This lack of relation appears at census zone level and is confirmed at sector level whose analysis does not show a correspondence between a social organization and a spatial organization. Only certain types of very marginal households (students, tenants) tend to group geographically. Globally, the elements structuring the space of the city at Sector level are mostly defined by the date of installation in Ouagadougou and on the parcel, dividing the periphery according to the contrast between the old districts in the north and the more recent population districts in the south.

In conclusion, we have focused on following changes in structures by simultaneously considering the steps of a scale system over four levels from household and parcel (socio-economic structure, urban integration and housing access mode levels) passing by the census zone (relation between urban fabric restructuring and residential structures) to the city sectors (planning operations levels). We have limited ourselves to a brief analysis of the main structures (provided by the main factor plans) of the scale system. It is nevertheless enough to bring out what is significant in the Ouagadougou space structuring of 1987: the major socio-demographical factors structuring small-scale levels, characterized by life cycle and intra-urban mobility have disappeared at the sector level. They are no longer discriminatory for the characterization of this sectioning of the city and only certain specific profiles (students-tenants, city natives) are concentrated on very specific spaces; a large population dispersion emerges over all of the Ouagadougou periphery, which does not get organized either spatially or in a particular housing operation. The analysis between two levels emphasizes the heterogenity and the dispersion elements at the heart of spatial units defining the higher level and makes it possible to understand the way in which spatial structures are modified. It remains interesting, even when structures are not modified much from one level to the other, since it reveals the nuances in intraclass structures.

2.4. Integration of the different levels

We now assume that the units of the analyzed level are "thrown" into different contexts or spatial environments playing a constraining role for the studied phenomenon, as was introduced in section 2.2. These contexts define higher level entities and the units of the considered level are analyzed as "all things being equal as for their context". We are using a confirmatory approach. Besides the methods used, it is the conceptual frame that changes, enabling a study focused of the event and the consideration of assumptions associated with the spatial organization of the researched units. Moreover, the approach in terms of context will consider a larger diversity of levels, integrating the territorial as well as dimension levels by proximity relations. In this section, we present local analysis descriptive methods, which are somewhat similar in their mechanisms to the method previously presented, except for what applies to quantitative attributes.

We will attempt to illustrate these methods from a simple example, using univariate information describing European regions. The generalization of this approach with a multivariate table will help us present the analysis of local differences developed by Benali and Escofier [BEN 88]. Finally, other local analysis methods will be described.

2.4.1. *The scale: a set of territorial and spatial references*

The example presented here comes from a contribution to the question of European integration from the standpoint of spatial disparities in the distribution of the level of wealth in the European regions[12] [COL 00]. The spatial heterogenity in

---

[12] The studied area here and called "Europe" is made up of the 15 countries that were EU members before 2004 with Switzerland and Norway, in total 17 countries. The observation level is that of regions NUTS3 (level 3 of the Nomenclature of Territorial Units for Statistics) made up of 452 regions. Wealth is measured by GDP per capita in 1990 (source: IGEAT, ULB). For reasons of changes in divisions, Portugal and the former East Germany are not represented.

the group of regions (I), is explained in terms of spatial continuity and discontinuity based on three levels, thus defining three relevant study grids (see Figure 2.8):

– the level of Europe (E) because it is inherent to the question as being the territory in construction, for which issues of heterogenity are raised;

– the national level (C) because the integration process must try to erase differentiations due to national affiliations and for which national hierarchies are a vector. The national level sets a partition of regions; we name it $C = \{c = 1, ..., m\}$ where n_c indicates the number of regions in I belonging to country c;

– the level of neighboring regions (V) for adjacency because the neighborhood represents another integration vector that can be based on an increase in local interactions between regions. The level of neighborhoods does not form a partition of regions, instead $V = \{V_i, i = 1, ..., n\}$ and n_{vi} indicates the number of regions of neighborhood V_i.

Each unit from one of the higher three levels (V, C or E) defines a context for one or several European regions of I. The heterogenity of the division of wealth will be successively evaluated in relation to each of the "local references": the European Union, the countries and finally the neighboring regions.

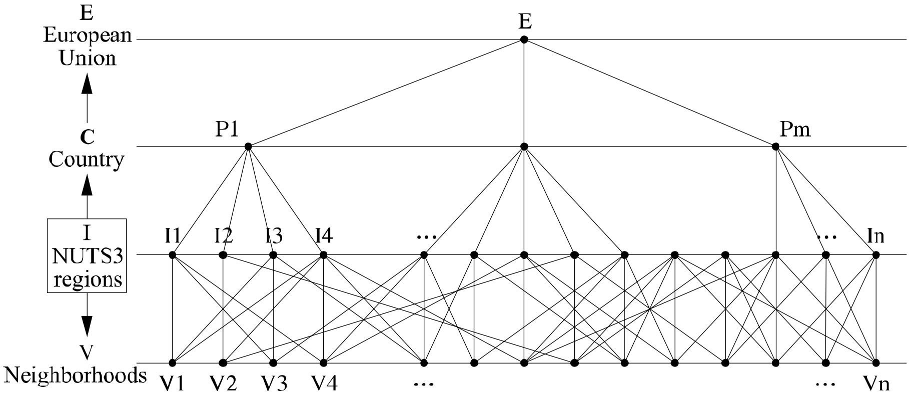


Figure 2.8. *Hierarchical structure of data*

X indicates the added value per capita in 1990 measured at regional level. We build an indicator that measures the differentiation between a region and its associated context, for each of the three levels:

$$\partial_i^N = X_i - X_k \qquad [2.9]$$

where N represents one of the three levels considered (E, C or V) and k is the index of the N class associated with region i.

Using the same matrix notations as those introduced in section 2.2.2, this transformation is written as: $\Delta^N = X - D^N G^N X$, where:

– X is the matrix $(n,1)$ for wealth data for each region;

– G^N is the matrix (n,n) of the proximity graph defined for I by N;

– D^N is the diagonal matrix (n,n) of the reverse of line G^N weight.

When N designates E, the European level, the indicator is the result of the use of average global values to center the data. On the other hand, if N designates one or the other national levels (C) or neighborhoods (V), the standardizing operation is no longer homogenous for all the studied population, but on unit subsets with the same context and we then talk about local differences.

The analysis of the first indicator ∂^E identifies the diversity and structure of the interregional distribution of wealth in Europe, with leading and slower regions (see Figure 2.9a). The analysis of ∂^C proposes a second study grid putting the European regions back in their national context. The wealth of each region is compared to the wealth of its country: two regions will be similar if they have the same position within their national system. We can then move on from general trends, mainly linked to national affiliations, to focus on intranational differentiations. Figure 2.9b illustrates this distribution and characterizes the control points of a national hierarchies-based homogenization. Finally, a third explanation is given by indicator ∂^V. It is based on the relevance of another local level, neighborhoods. The wealth of each unit is expressed in terms of deviation from the average wealth of the region which it constitutes with the regions surrounding it. The mapping of this last indicator (see Figure 2.9c) highlights local maxima and minima, the former appearing as central nodes of local networks based on an assumption of diffusion by proximity. This neighborhood level, which is intermediate between the regional and national levels, is not part of the administrative hierarchy. A large part of the entities in this level are transborder. This is the major significance of levels defined by proximity graphs: they enable an integration of space outside of the strictly territorial frame and thus a more prospective vision of the distribution.

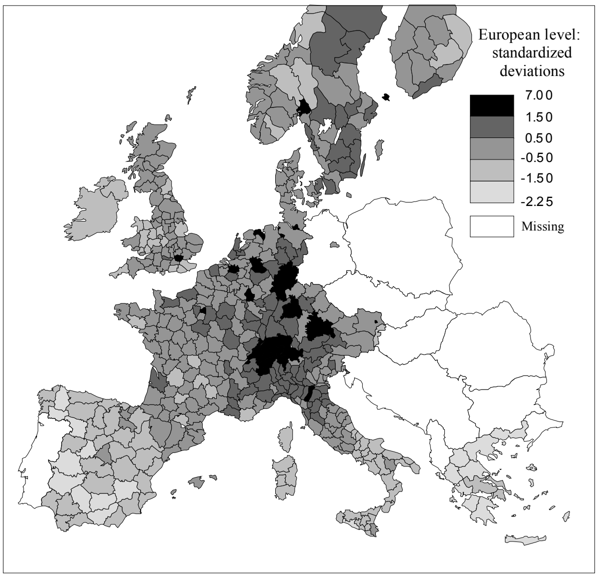


Figure 2.9a. *Wealth disparities: deviations at European level*

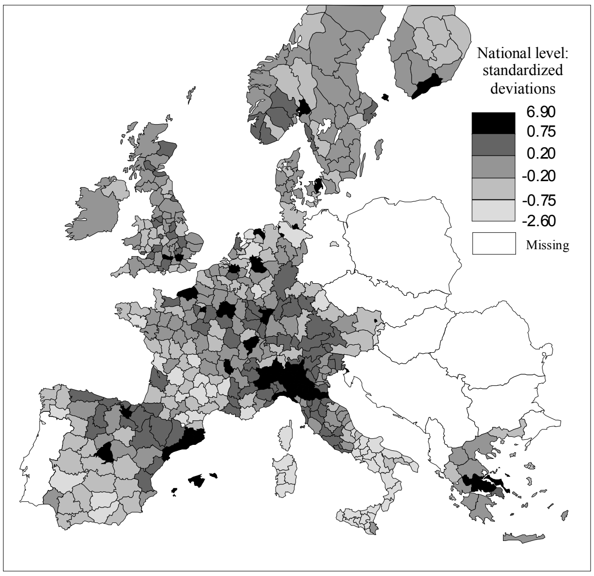


Figure 2.9b. *Wealth disparities: deviations at national level*

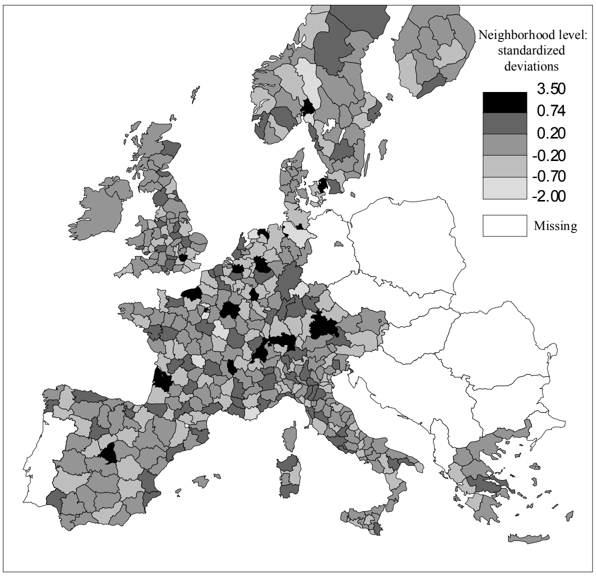


Figure 2.9c. *Wealth disparities: deviations at neighborhood level*

This example not only illustrates a possible use of maps, the indicators shown have been built from considerations on different relevant scales and levels in order to be able to make hypotheses on the components of the integration process. The map is the indispensable support of this information, but the spatial or territorial structure of the system is integrated from the initial stages of the process and makes it possible to understand the spatial character of the integration.

2.4.2. *The analysis of local differences*

This approach is similar to a multivariate table. When the higher level N is a partition of I (as is the case with countries), the method resembles that of an analysis of intraclass differences that we have already discussed in section 2.3 in the case of qualitative variables. In a more general case where the higher level is not a partition of I (for example, contiguous neighborhood levels), it is called analysis of *local differences*. We will present an application of this analysis by going back to the previous example, in which the European regions are described by the distribution of the added value per capita in nine economic sectors[13], let X be the geographical information table describing units i of I by l variables of general term X_{ij}, with $i = 1, ..., n$ and $j = 1, \ldots, l$. X_{ij} indicating the measure for unit i of the j^{th} variable. The variables of matrix X of size (n,l) are centered and reduced. The objective is the description of the local heterogenity of the units, calculating the ratio of global variability to any level N defining m classes k on I.

The factor analysis of local differences proceeds in a principal component analysis (PCA), based on the table of local differences where each individual is described by the difference with the average point of the class Q to which it belongs. Using the notations introduced in section 2.2.3, the matrix form of the analyzed table that we will note by Δ^Q is:

$$\Delta^Q = X - D^Q G^Q X \qquad [2.10]$$

Q explains one of the three levels: E, C or V. Here $\Delta^E = X$. Two more tables are constructed; one is relative to local differences associated with countries Δ^C and the other is relative to local differences associated with neighborhoods Δ^V. Table X is centered as Δ^P is, because P forms a partition of I, whereas Δ^V is not centered. Neither one of the two is reduced. In the case where the analysis of the main components is done on the non-reduced table Δ^P, the results of the analysis are directly comparable with those of table X, since the inertia of table X is decomposed according to partition C in two complementary parts: the interclass inertia and the intraclass inertia which is the inertia linked to table Δ^C. This is not the case with the analysis of table Δ^V. If it is preferable that table Δ^V is centered beforehand in order to apply the PCA, whether it is reduced or not does not make the results of the analysis similar to those in table X because V does not represent a partition of I[14].

The PCA of table $\Delta^E = X$ emphasizes the major principles of the economic specialization differentiations between the European regions when they are considered as a whole set. The three first factors summarize 60% of these differentiations as follows: the first factor (30%) opposes the agricultural sector to the industrial (equipment or construction), commerce and services sectors: the main differentiation between the regions is either an over-representation in one of the first sector or in one of the second sector. The next two factors show the same importance, each making up about 15% of total differentiations. The first opposes the regions in which agricultural and industrial sectors, as well as the consumer goods sector, are relatively important to the regions in which these sectors are under

---

[13] The nine sectors are: agriculture, food industry, construction, intermediate goods, capital goods, consumer goods, transport and communication, commerce and finally services including banks, insurance and goods and services.

[14] These notions are explained in more detail in [BEN 90].

represented. The second concerns the transport and communications sectors to intermediate, consumer and capital goods sectors.

The PCA of table Δ^C enables us to analyze the entire European regions based on the sectoral structure of the intranational differentiations of economic specializations. The first three factors of this analysis also explain 60% of the differentiations of regions relative to the "average profile" of the country to which they belong. The main trends in national differentiations are almost identical to the differentiations observed in all of Europe. The structure observed at the European level is reproduced on average for each country. The similarities here should be interpreted relatively to the national average profiles, that is, two regions are similar in the economic sector space if they deviate from their national average profile in the same direction. Countries then differenciate depending on the intensity of these contrasts.

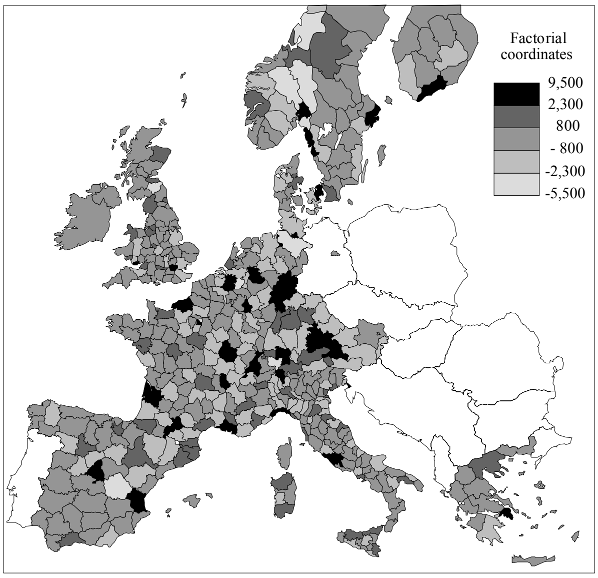


Figure 2.10. *Differentiation at neighborhood level*

Finally, the analysis of table Δ^V explains the major differentiation factors between the regions and their neighborhoods. As with the first two analyses, the three first factors explain more than 55% of all the differences between each region and its neighborhood. The first factor is very similar to that of the previous analyses, thus confirming that, even on a local level, the biggest differentiation (27%)

between the units and their neighborhoods is on average due to differences in agricultural specializations on the one hand (rural regions) and to commerce and services on the other hand (urban regions). The next two factors show comparisons and associations slightly different from the first two analyses. The second factor opposes regions that have locally high added values in relation to the agricultural and food sectors, or with construction, for regions that have a local advantage in intermediate or capital goods sectors. The third factor associates construction and intermediate goods sectors to which the consumer goods sector is opposed.

Figure 2.10 illustrates the distribution of the first local differentiation factor associated with neighborhoods. The distribution is less "regionalized" in certain parts of Europe, thus illustrating the local differentiations of heterogenity. It makes it possible to identify local specializations to understand their organization in terms of gradients and discontinuities.

2.4.3. *Other local analysis methods*

Other methods have been developed with the same objective: integrating the existing structure over all the studied units in order to emphasize spatial relations. The oldest method is attributed to Lebart [LEB 69, LEB 84] who, from Geary's [GEA 54] spatial autocorrelation formulation, proposed local factor analyses based on the local relations between variables by using the local variance-covariance matrix (W^L) associated with X and G, the proximity graph instead of the variance-covariance W matrix associated with table X. W and W^L are given by: $W = X'X/n$ and $W^L = [X' (D^{-1} - G) X]/m$, where m represents twice the number of related unit pairs in G.

In Lebart's local analysis, the individuals are used with a weight proportional to the size of the neighborhood associated with them. Benali shows that this method is similar to the one of local differences if the classes have the same weight [BEN 90]. It is important to note, however, that it cannot be implemented with a traditional PCA program.

An analysis similar to the local analysis as proposed by Lebart is one of a table where the rows do not describe the units but describe the pairs of related units instead. Grasland [GRA 97] has analyzed the demographic discontinuities of European regions by applying a PCA to a table associating to each contiguous pair of regions the age structure differences between these regions, an analysis which he calls "analysis of local differentiations". The author recalls that in the analysis of a table, the variance of the variables reached is proportional to the local variance of the variable defined in *I*.

Finally, on the basis of the transformations presented in sections 2.4.1 and 2.4.2, we can consider the construction of tables integrating simultaneously several levels of contexts associated with different operations. Thus described, an entity classification will group the entities relative to these different contexts. In [COL 00] or in [SAN 00] we will find this type of applications where the units studied are put in relation to several levels of contexts, including neighboring units, but also of contexts defined by the local urban network.

We have illustrated how from a single table of data, without changing the goal of the study, several study grids can be built, since several images from different levels successively play the role of references. The differentiation factors emerging from this type of analysis show the structure associated with each of these referentials, their comparison emphasizing the permanencies or the changes between referentials.

2.5. Multilevel models[15]

The two previous examples differ in their conceptual approach from the notion of hierarchy of levels (aggregation effect and contextual effect), although they are both based on descriptive analysis methods. The explanatory approaches have produced methodological developments enabling a modeling of differentiations between individuals which takes into account the hierarchical structure of the data: *multilevel statistical models* [GOL 95], also more generally termed *mixed models*. These methods have been developed within microscopic analyses to move beyond one of the big “dilemmas” often experienced by human sciences modeling specialists: the choice of the most relevant level, microscopic or aggregated (see Chapter 6). They offer the possibility of analyzing the characteristics of a set of individuals keeping in mind the differentiations related to the environment or the context in which they evolve. There are many applications of these methods in the following fields:

– in education sciences, where school results are modeled according to individual student characteristics and also by the characteristics of the class and the school [BRE 97, MOB 02] or the role of the shared family context in understanding differential parental treatment of children [JEN 03];

– in epidemiology, where the probability of the prevalence of a pathology is a function of the inherent characteristics of the individuals and also of their environment [DUN 97, JON 95];

– in voting behavior analysis, where the influence of context has been quite well demonstrated [JON 97];

---

[15] Section written by Hélène MATHIAN and France GUERIN-PACE (researcher at INED).

– in migration analysis, where the probabilities for migrating from one region to another depend not only on the social characteristics of the people but also on the characteristics of the origin and destination regions [COU 03].

All these applications have an obvious geographical dimension. The observation level is the individual level and the characteristics of places are introduced as explanatory factors.

In this section we propose to compare this type of method with the previous applications. More than just a description of the methodology itself, for which we can reference specialized studies and books [COU 03, GOL 95], we are suggesting an introduction of the method in a simple geographical illustration. The objects studied are not individuals but places that can be described by inherent or aggregated attributes. The multilevel models have been developed and applied in several areas: regression, logistical model, biographical models, etc. We are presenting the method in the context of a regression model.

2.5.1. *Contextual effects and regression models*

We attempt to model the population growth of a group of French towns over a period T_1 (1990-1999) according to their growth in the previous period T_0 (1982-1990). We are considering 4,500 municipalities[16] spread over 17 departments[17] in the South of France, where I is the set of municipalities (i), and D is the group of counties (d); D is a partition of I.

For simplicity, in this case only the growth of the previous period is used to explain the growth of the current period. We could introduce other variables that characterize the towns: the number of jobs in the town, the distance to the closest city of a given size, the proportion of youth, etc., but we have chosen to look at a simple regression.

Analyzing the general form of the relation between both growth rates over the group I of studied counties means testing the model:

$$\forall\, i \in I, \qquad Y_i = aX_i + b + e_i \qquad [2.11]$$

---

[16] Only municipalities with more than 100 residents in 1982 have been retained, thus making a total of 4,469 municipalities.

[17] Departments constitute the zoning called NUTS3 at the European level, such as counties in the UK.

where X_i and Y_i are respectively the initial growth rate (period T_0) and the final growth rate (period T_1) of county *i*.

On the group of studied counties, this model shows a final growth rate variance of 18% (Prob < 0.0001) and is noted by: $Y = 0.306\,X + 0.34$, representing the average cumulative growth effect.

In such a model, the group of municipalities is considered as forming a homogenous group relative to the analyzed relation. There are numerous factors from higher levels differentiating the municipalities within this relation, such as, for example, local dynamics, one of the components of which is urban dynamics. Indeed, the 17 departments constituting the studied region show a large diversity of situations with regard to the demographic dynamic, which goes from highly rural departments with high population decrease, such as Lozère and Aveyron, to departments with a high demographic growth, concentrated around developing cities such as Haute-Garonne, Hérault and Bouches-du-Rhône. We can assume that a contextual effect, which is defined here by the departmental affiliation[18], would influence the shape of the relation.

This is equivalent to the modeling of the relation by introducing an interaction between the growth of municipalities and departments. Figure 2.11, based on the example of Jones [JON 97, p. 21], theoretically illustrates different types of interactions between the explanatory variable (X) and the contextual factor for the explanation of the dependent variable (Y). Applied to our example, Figure 2.11a would illustrate the relation between initial and final growth rate over all the municipalities studied, whereas Figures 2.11b to 2.11e illustrate the different possible configurations if this relation was modeled county by county: each straight line is associated with a municipality, whereas the bold straight line represents the general relation. This relation is written as:

$$Y_{id} = a_d X_{id} + b_d + e_i \qquad [2.12]$$

where $i \in I$ represents a municipality, and d the department of D to which i belongs.

Figure 2.11a illustrates a positive general relation between variation rates of periods T_0 and T_1 showing temporal auto-correlation of variation rates for the two intercensal periods. On average, the population variation of the municipalities between 1982 and 1990 tends to continue between 1990 and 1999.

---

[18] Although the urban growth context is not limited to department boundaries, we choose to use these limits to define "simple" contexts forming a partition of space.

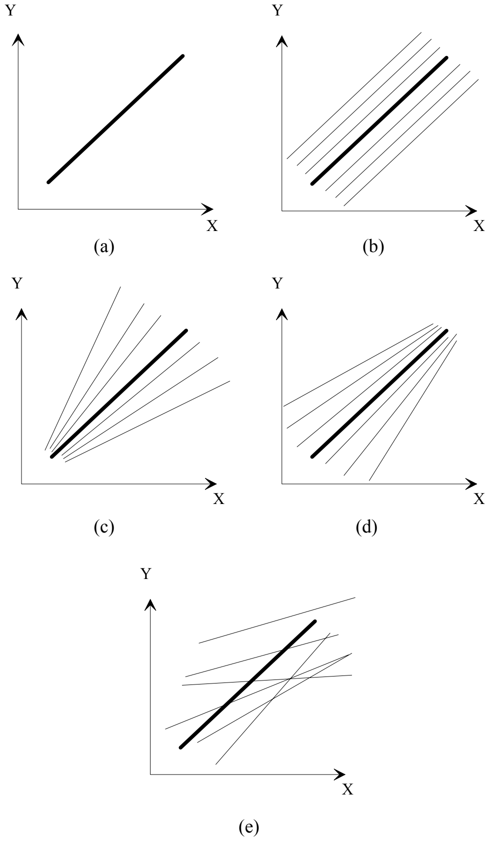


Figure 2.11. *General relation and context effect (based on [JON 97])*

If we now analyze this relation at department level, several cases are possible:

– Figure 2.11b: whatever the department, the form of the relation is identical; we obtain a series of parallel straight lines. The difference between one department and another is the growth intensity whose average rate will be at different levels in each county;

– Figure 2.11c: the form of the relation differs according to the departments. Straight lines are arranged in an array and are different mainly by their slopes. The graph illustrates a differentiation arranged according to two components; interdepartment differentiation and intradepartment differentiation. If, from one department to another, the average values do not change much compared to the previous graph, this differentiation is weaker between classes of municipalities which, within each department, have experienced relatively low growth rates during

T_0, whereas it will be stronger for those experiencing high levels. In fact, intradepartment differentiations vary from one department to another. For those departments that have the highest average rate, growth was higher in the municipalities with the highest initial growth rate, thus creating strong intradepartment differentiations, whereas for the less dynamic departments, the growth has decreased for these same municipalities, reducing the disparities between them;

– Figure 2.11d: we have here the opposite situation from the previous figure. The interdepartment differentiation is in fact stronger for classes with low initial rates. The intradepartment differentiation decreases for the departments with high final average rates, whereas it increases for the other departments;

– Figure 2.11e: this figure illustrates a more complex interaction between growth and places. In certain departments, initial growth is positively linked to final growth, whereas in others there is little influence. We could even consider counties where this relation is negative.

This is a schematic representation. Table 2.1 presents the results of the 17 regressions calculated by county. The simultaneous representation of the 17 adjustment straight lines could remind us of Figure 2.11e.

The consideration of contextual effects by calculating as many regressions as there are contexts or higher level classes raises two concerns: the first, which is practical, concerns the size of sub-populations; we rarely have a sufficiently high number of observations for the relations to be significant (which is not the case here). The second concern, which is more theoretical, is to consider that the classes formed by the higher level (the departments in this case) constitute independent groups.

A solution consists of starting with the general model [2.11] and, by integrating the higher level via contextual variables, which describe the higher level and take differentiated values according to the classes. Realistically, we can imagine introducing in the regression model either aggregated variables shown in the form of averages (for example, average department variation rate), or proportions (proportion of urban population in the department), or even any other variable defined at this level and presumed to have an effect on the town growth. The model then becomes[19]:

$$X_{id} = a_1 X_{id}^{1} + a_2 X_{id}^{2} + b + e_i \qquad [2.13]$$

where X_{id}^{1} is the initial growth rate of town *i* of county *d* and $X_{id}^{2}=X_{d}^{2}$ is the growth rate of county *d* of *D* to which the town *i* belongs.

---

[19] In specialized works, such a model is called a contextual model, contrary to multilevel models.

Departments	Degree of freedom	Constant	Coefficient	Part of explained variance R^2	Signifi-cativity
Ardèche	305	0.344	0.271	0.147	0.0001
Ariège	204	0.346	0.093	0.026	0.0203
Aude	295	0.371	0.130	0.041	0.0004
Aveyron	293	-0.227	0.232	0.077	0.0001
Bouches-du-Rhône	117	0.761	0.236	0.182	0.0001
Drôme	261	0.791	0.095	0.020	0.0221
Gard	322	0.685	0.270	0.149	0.0001
Haute-Garonne	378	0.621	0.419	0.248	0.0001
Gers	440	0.014	0.281	0.117	0.0001
Hérault	300	0.936	0.331	0.248	0.0001
Lot	307	0.312	0.123	0.019	0.0145
Lozère	148	0.118	0.198	0.064	0.0018
Hautes-Pyrénées	299	-0.059	0.168	0.053	0.0001
Pyrénées-Orientales	166	0.766	0.187	0.090	0.0001
Tarn	287	0.095	0.268	0.146	0.0001
Tarn-et-Garonne	177	0.192	0.098	0.017	0.0809
Vaucluse	136	0.952	0.124	0.030	0.0429

Table 2.1. *Parameters associated with the analysis per department*

The introduction of a contextual variable improves the quality of the estimation ($R^2 = 0.22$). Table 2.2 presents the associated parameter values. The intrinsic effect of the initial growth rate of the town decreases (coefficient of 0.262), whereas the contextual effect (growth rate of the department to which the municipality belongs) has a significant positive effect (coefficient of 0.433). However, with such a formulation, we consider that the effect of the contextual variable (dynamic demographics of the county) is the same for all the towns of the department. However, as we have seen previously (see Figure 2.11), it can be different for each department, because it can be affected by characteristics inherent to the departments, such as the number of cities, the distance between cities, the hierarchical form of the population, accessibility, etc.

2.5.2. *Multilevel modeling*

Multilevel models make the assumption that to individual variations (towns in this case) surrounding the general form of the relation, variations linked to context effects are added and modify the intensity and the form of the relation (slope and/or constant value) (see Figure 2.11).

We can certainly modelize these variations in a traditional way by introducing affiliation to the counties through dichotomous variables, each variable being associated with a department. We obtain different slopes and/or constant parameters according to the counties, which would undoubtedly improve the model. However, this advantage is limited by the adjustment method of least squares used for the resolution in which the residues are supposed to be independent from each other. This assumption, as noted by Bressoux *et al.* [BRE 97], denies the effect related to context and the fact that the growth rate in municipalities can also depend on the growth rate of all the municipalities belonging to the same department.

Multilevel models make it possible to integrate this dependence which reflects the differentiating effect of the environment, by introducing random effects at the department level. These effects are considered as coming from a random draw among a larger group of departments. The parameters become random variables linked to departments. The multilevel model is written as:

$$\forall\ i \in d \text{ and } d \in D \qquad Y_{id} = a_d X_{id} + b_d + e_{id} \qquad [2.14]$$

where a_d and b_d are random variables of average a and b respectively.

There is the relation $a_d = a + v_d$ and $b_d = b + u_d$, in which u_d and v_d characterize department variations around these averages.

Several procedures generalizing the least squares method have been developed for the estimation of all of these parameters [GOL 95]: fixed effects (*a* and *b*) and the random effects. u_d, v_d and e_{id} are random variables of zero average and of variance $\sigma^2_u = \text{var}(u_d)$, $\sigma^2_v = \text{var}(v_d)$, $\sigma^2_e = \text{var}(e_{id})$, and of covariance $\sigma^2_{uv} = \text{cov}(u_d;v_d)$.

Transposed in our example, the residual variance is broken down in a municipal variance (σ^2_e) and an interdepartment variance associated on the one hand with constants (σ^2_u) and on the other hand with slopes (σ^2_v).

Parameters of the model Estimations* (standard errors)	**General (model [2.11])**	**Contextual (model [2.13])**	**Multilevel (model [2.14])**
Fixed effects			
Constant	**0.340** (0.021)	**0.097** (0.027)	**0.397** (0.084)
municipality level:			
Initial growth rate (1982-1990)	**0.306** (0.010)	**0.262** (0.010)	**0.218** (0.023)
department level:			
Average growth rate of department between 1982 and 1990		**0.433** (0.03)	
Random effects			
Municipality effect:			
Individual variance s^2_e	1.796 (0.038)	1.718 (0.038)	1.650 (0.035)
Department level:			
Constant variance s^2_u			**0.110** (0.041)
Slope variance s^2_v			**0.007** (0.003)
Constant/slope covariance s^2_{uV}			0.005 (0.008)
Statistics			
Part of the explained variance (R^2)	0.18	0.22	
Maximum likelihood (-2Log(V))	15,299	15,098	14,994

** In bold, significant coefficients.*

Table 2.2. *Comparison of the different models: fixed and random effects*

In Table 2.2, we see the estimations of the parameters of the different models applied to the municipalities in the 17 departments. The degree of significance of the model is measured by comparison from the maximum likelihood statistic (-2Log(V)) and that of the coefficients (in bold) is read from standard errors of the estimations (in parentheses). The multilevel formulation improves the explanatory relevance of the model. Even more so than with the contextual model, the effect of the initial growth rate is decreased in relation to the general model, confirming that a part of this effect is in reality linked to context differences. The "department" effect is significant. We notice a strong variance of constants, showing what certain authors call the "efficiency" differences between the groups [BRE 97]. This term, used in education sciences, is not very appropriate for our example in which the counties do not have performance objectives even if, all initial growth rates being equal, certain departments have higher growth levels. Slope variance also has a significant value. By symmetry, the same authors associate the notion of "equity" with this variance. In fact, different heterogenities of the final rates are associated with slope differences. In certain departments, growth is concentrated in a few municipalities with better local conditions, thus creating strong differences, whereas in others the situation is much more "egalitarian", with the rate variability being minimal. On the other hand, the two effects do not converge (insignificant covariance). A significant value should have explained the presence of a cumulative effect in which the higher the initial growth rate, the higher the probability that growth would happen to municipalities initially in a period of growth.

The multilevel formulation of such a model can then be enhanced with the addition of other variables, which are defined at municipality or department level. We can then integrate a certain number of characteristics which make it possible to test the effect of the relative position of the cities within a department. For example, would a hierarchical departmental network of cities be more efficient? What would the effect of spatial distribution of cities be? These questions call for characteristics inherent to municipalities (location in relation to the network of cities) as well as aggregated characteristics (form of the network of cities). We have deliberately shown here a simple example, using two nested levels. However, these models can integrate much more levels, nested or not, and help to test the fixed and random effects associated with each level. In our example, we could introduce the level of "urban areas"[20] and test the differentiations associated with this space hierarchization for the analysis of growth. We could also be tempted to introduce contextual effects linked to neighborhood, such as those defined in section 2.4, but the level of neighborhoods that do not form a partition of space cannot be directly introduced. It is therefore necessary to use a matrix formulation of the relation, like that explained in section 2.2.2.

---

20 In 1990, the INSEE defined local zoning in urban and rural areas, based on the notion of "urban influence".

These methods are very useful as long as they can process complex interactions between effects associated with different simultaneous analysis levels. Its application to places (aggregates) rather than to individuals remains, for now, very marginal. As is the case with any other type of modeling, primary level residual variance generally remains strong, summarizing a group of diverse local factors that we cannot explain with simple quantitative measures. These are the famous deviations in relation to the models that most often result in individual initiatives and that send us back to a larger debate: what are the relevant modeling levels?

2.6. Conclusion

These different applications show the interest of the adaptation of multidimensional statistical methods to take into account the spatial data structure. Space is considered throughout all the relations that structure its different components. The formalization of these relations, proposed in this chapter, is based on proximity (neighborhood, affiliation), which is a very flexible and generic notion for many situations in the processing of space. This formalization can integrate and explain the links between entities of the same level or of different levels, nested or not. The description of spatial structures associated with the different levels, their comparison and the consideration of their variability constitute genuine multilevel analysis systems.

Several methods are developed and consider this specificity to confirm the spatial structure and to measure its effects. They are the answer to a real necessity of comparing statistical distributions and territorial and spatial structures, statistical heterogenity and discontinuity in space. Local analysis methods and multilevel methods appear complementary: they cover all the approaches that we have described. The local analysis methods presented here are appropriate when there are *a priori* assumptions or information on observation or individual pairs, which is the case with geographical information. They have the advantage of being easily implemented with traditional statistical tools, as long as we have ways to formalize the relations between the units that we attempt to develop or to integrate in the analysis by proximity or partition graphs. The multilevel model, on the other hand, uses specific resolution methods which are rarely integrated in traditional statistics software[21]. Nevertheless, the benefit of simultaneously integrating different aggregation levels resides in the possibility of attributing the residual variance of classical multiple regressions to each level. The feature of its application in geography, which may constitute one of its limits, comes from the fact that the microgeographical level is rarely that of individuals, but is already an aggregation

---

[21] We can name MLWIN, HLM, AML software packages or the procedure MIXED of SAS.

level. However, the diversity of individual behaviors must be returned, one way or another, in aggregated data modeling.

We have resolutely chosen a methodological and statistical approach. These approaches, which are used jointly with specific methods developed by geographers or other specialists of spatial phenomena, perfectly integrate in a spatial analysis approach in an attempt to advance about the connection between phenomena operating at different levels in space.

2.7. Bibliography

[BEN 88] BENALI H., ESCOFIER B. "Smooth factorial analysis and factorial analysis of local differences", in *Analysis of Multiway Data Matrices*, R. Coppi and S. Bolasco (eds.), Amsterdam, 1988.

[BEN 00] BENALI H., "L'analyse des correspondances multiples lissées et l'analyse des correspondances multiples des différences locales", in *L'analyse des correspondances et des techniques connexes: approches nouvelles pour l'analyse statistique des données*, Springer, Mathématiques & Applications vol. 32, Berlin, p. 121-143, 2000.

[BENZ 73] BENZECRI J.P., *L'analyse des données. Tome 2: l'analyse des correspondances*, Dunod, 1973.

[BRE 97] BRESSOUX P., COUSTERE P., LEROY-AUDOUIN C., "Les modèles multiniveau dans l'analyse écologique: le cas de la recherche en education", *Revue française de sociologie*, XXXVIII, p. 67-96, 1997.

[BRU 92] BRUNET R., FERRAS R., THÉRY H., *Les mots de la géographie*, RECLUS-La documentation Française, Collection Dynamiques du territoire, Montpellier-Paris, 1992.

[CHA 95] CHARRE J., *Statistique et territoire*, RECLUS, Collection Espaces mode d'emploi, Montpellier, 1995.

[CLI 73] CLIFF A., ORD J.K., *Spatial Autocorrelation*, Pion, Paris, 1973.

[COL 00] COLL J.L., GRASLAND C., PUMAIN D., SAINT-JULIEN T., *Bilans et perspectives pour la constitution d'un Observatoire de Recherche pour l'Aménagement du Territoire en Europe*, DATAR, Paris, 2000.

[COU 03] COURGEAU D., *Methodology and Epistemology of Multilevel Analysis Approaches from Different Social Sciences*, Kluwer Academic Publishing, Dordrecht, 2003.

[DAU 94] D'AUBIGNY C., D'AUBIGNY G., "Agrégation spatiale et résumés statistiques", *Revue internationale de Géomatique*, vol. 4, no. 3/4, p. 307-336, Hermès, 1994.

[DUM 97] DUMOLARD P., "Validation d'un découpage territorial", in *Les découpages du territoire*, INSEE méthodes, Dixièmes entretiens Jacques Cartier, p. 59-69, 1997.

[DUN 97] DUNCAN C., "Applying mixed multivariate multilevel models in geographical research", in *Places and People: Multilevel Modelling in Geographical Research*, Westert G.P., Verhoeff R.N. (eds.), The Royal Dutch Geographical Society, p. 9-20, 1997.

[DUP 94] DUPÉRIER E., "Changement d'échelle et géographie des élections: le vote des villes françaises en 1988", *L'Espace Géographique*, vol. XXIII, no. 2, p. 167-174, 1994.

[DUR 99] DURAND-DASTÈS F., "L'explication en géographie" (1974, unpublished), in *Systèmes et modèles*, CD Rom, UMR Géographie-cités, Paris, 1999.

[ESC 87] ESCOFIER B., "Analyse des correspondances multiples conditionnelles", *Data Analysis and Informatices*, North-Holland, p. 13-22, 1987.

[FOT 91] FOTHERINGHAM A.S., WONG D.W.S., "The modifiable areal unit problem in multivariate statistical analysis", *Environment and Planning A*, p. 1025-1044, 1991.

[GEA 54] GEARY R.C., "The contiguity ratio and statistical mapping, the incorporated statistician", vol. 5, no. 3, p. 155-145, 1954.

[GOL 95] GOLDSTEIN H., *Multilevel Statistical Models*, Arnold, New York, 1995.

[GRA 97] GRASLAND C., "L'analyse des discontinuités territoriales: l'exemple de la structure par âge dans les régions européennes", *L'espace Géographique*, 4, p. 1-18, 1997.

[GRA 01] GRASLAND C., MATHIAN H., VINCENT J.M., "Multiscalar Analysis and Map Generalisation of Discrete Social Phenomena: Statistical Problems and Political Consequences", *UNECE Statistical Journal*, p. 157-188, 2001.

[GRO 97] GROENEWEGEN P., "Dealing with micro-macro relations: a heuristic approach with examples from health services research", in *Places and People: Multilevel Modelling in Geographical Research*, Westert G.P., Verhoeff R.N. (eds.), The Royal Dutch Geographical Society, p. 100-117, 1997.

[HAG 65] HAGGETT P., *Locational Analysis in Geography*, London, E. Arnold, Paris, 1965.

[JEN 03] JENKINS, J.M., RASBASH J., "The role of the shared family context in differential parenting", *Developmental Psychology*, 39(1), p. 99-113, 2003.

[JON 95] JONES K., DUNCAN C., "Individuals and their ecologies: analysing the geography of chronic illness within a multilevel modelling framework", *Health and Places*, vol 1, no. 1, p. 27-40, 1995.

[JON 97] JONES K., "Multilevel approaches to modelling contextuality: from nuisance to substance in the analysis of voting behaviour", in *Places and People: Multilevel Modelling in Geographical Research*, Westert G.P., Verhoeff R.N. (eds.), The Royal Dutch Geographical Society, p. 19-43, 1997.

[LEB 69] LEBART L., "Analyse statistique de la contiguïté", *Publ. Inst. Stat. Univ. Paris*, XVIII, p. 81-112, 1969.

[LEB 84] LEBART L., "Correspondance analysis of graph structure", *Bulletin technique du centre de statistique et Informatique appliquée*, 3, p. 5-19, 1984.

[LEB 06] LEBART L., PIRON M., MORINEAU A., *Statistique exploratoire multidimensionnelle : visualisation et inférence en fouilles de données*, Dunod, Paris, 2006.

[LEBR 92] LEBRIS E., JAGLIN S., MARI A., OSMONT A., OUATTARA A., OUEDRAOGO J.-P., PIRON M., SANOU L., Les enjeux des extensions urbaines à Ouagadougou, Rapport Orstom-IFU-MRT, 1992.

[MAR 99] MARCEAU D., 1999, "The scale issue in the social and natural sciences", *Can. J. Remote Sensing*, 25: 347–356.

[MOB 02] MOBLEY C., "Community colleges and the school-to-work transition: a multilevel analysis", *Sociological Inquiry*, 72(2), p. 256-284, 2002.

[MUL 91] MULLON C. (ed.), *Transfert d'échelle*, Orstom Editions, 1991.

[OPE 84] OPENSHAW S., "The modifiable aerial unit problem", UK GeoBooks, *Concepts and Techniques in Modern Geography*, vol. 38, Norwich, 1984.

[OPE 96] OPENSHAW S., "Developing GIS-relevant zone-based spatial analysis methods", in *Spatial Analysis: Modelling in a GIS Environment*, Longley P., Batty M. (eds.), GeoInformation International, Cambridge, p. 55-73, 1996.

[PAV 94] PAVÉ A., *Modélisation en biologie et en écologie*, Aléas Editions, Lyon, 1994.

[PIR 92] PIRON M., *Analyse statistique d'un système d'échelles*, A.D.O.C. network, Memorandum no. 4, Orstom Editions, 1992.

[PIR 93] PIRON M., "Changer d'échelle", *L'Espace Géographique*, no. 2, p. 147-165, 1993.

[PUM 97] PUMAIN D., SAINT-JULIEN T., *Analyse spatiale (1)*, Armand Colin, Cursus Collection, Paris, 1997.

[RAY 96] RAYNAL L., DUMOLARD P., D'AUBIGNY G., WEBER C., RIGAUX P., SCOLL M., LARCENA D., "Gérer et générer des données spatiales hiérarchisées", *Revue internationale de géomatique*, Vol. 6, no. 4, Hermès, p. 365-382, 1996.

[REY 98] REYNOLDS H.D., The Modifiable Area Unit Problem: empirical analysis by statistical simulation, PhD Thesis, Toronto, http://www.badpets.net/Thesis/index.html, 1998.

[ROZ 89] ROZENBLAT C., "Les quartiers de Paris", *L'Espace Géographique*, p. 264-272, 1989.

[SAN 89] SANDERS L., *L'analyse des données en géographie*, Alidade RECLUS-La Documentation Française, Montpellier-Paris, 1989.

[SAN 99] SANDERS L., GAUTIER D., MATHIAN H., "Les concepts de système spatial et de dynamique: un essai de formalization", *Revue Internationale de Géomatique*, Vol. 9, no. 1, Hermès, p. 25-44, 1999.

[SAN 00] SANDERS L., MATHIAN H., "Cities, commutings and urban contexts", in *Integrated Urban Systems and Sustainability of Urban Life*, Ianos I., Pumain D., Racine J.-B. (eds.), Editura Technica, Bucharest, p. 49-66, 2000.

[WRI 96] WRIGLEY N., HOLT T., STEEL D., TRANMER M., "Analysing, modelling and resolving the ecological fallacy", in *Spatial Analysis: Modelling in a GIS Environment*, Longley P., Batty M. (eds.), GeoInformation International, Cambridge, p. 23-54, 1996.

Chapter 3

Location of Public Services: From Theory to Application

3.1. Introduction

Public services occupy an important place in modern societies. We are born in a hospital, we go to school, we benefit from transportation infrastructures and emergency services, we communicate by mail, telephone and Internet, we go to a city hall for anything to do with civil status, our household garbage is collected, we are buried in a cemetery, etc. One of the challenges facing contemporary society is to reach a correct adequacy between its members' needs and the services that they offer. This adequacy has several facets and concerns: the identification of needs, the selection of those that should be satisfied, the way in which service providers will be organized and controlled, the arbitrage between users, the funding of these activities, the impact on the physical and human environment, etc. All these issues open the door for large debates in political, social, economic, ethical, etc., areas.

A question of particular interest to the geographer throughout these discussions is the spatial adequacy between supply and demand, which means the choice of the location of the services and their geographical coverage. The goal of this chapter is to show how operations research methods can provide answers to these questions. The problem is prevalent in economically developed regions. Indeed, most of these regions are faced with huge budget (and debt) control problems that require, in principle, the best possible allocation of public money. Moreover, because of the emergence of supranational institutions (European Union, NAFTA, etc.), many

---

Chapter written by Dominique PEETERS and Isabelle THOMAS.

competences are to be redistributed among the different levels of authorities. The issue of the spatial organization of public services is also critical in developing countries where services are to be created or reformed so that they are as efficient as possible in order to serve their population, while coping with the often limited means that the governments of these countries have at their disposal as well as poor communication networks.

At the heart of this problem on the location of public services is an antagonism between, on the one hand, the setting up and operating costs of the facilities and, on the other hand, the access costs borne by the users [BEG 92a]. Only considering the implementation costs would result in creating the smallest number of facilities compatible with the general level of demand, and only taking into account access costs would imply a proliferation of establishments so that the system would be as close as possible to the largest number of users. In the first case, the solution would impose considerable transport costs, substantially reducing the collective use or the quality of the services under consideration. In the second case, investment and operating costs would be exorbitant and would entail an unbearable financial burden for the community. Hence, a compromise has to be found between those two extreme solutions.

A second important question is about *equity* between members of society with regard to access to services, as the distribution of trip length is almost always non-egalitarian. However, the search for an equitable solution could lead to a socially inefficient answer, because there might be locations that would be closer to all users of the service, but at the cost of different access conditions. For example, in order for all the members of a community to be at the same distance from a service, we can simply locate this service at an infinite distance, thus making the service unavailable for everyone! It is therefore better to give up the search for a "just" location, in order to focus on an "equitable" location by maximizing the welfare of the worst-off user (this is the criteria advocated by the philosopher Rawls [RAW 71]). However, even the quest for a larger equity almost always leads to higher costs for the community. The question must therefore be discussed *of a balance between system efficiency and geographical equity for users as well as for the entire community*.

In practice, location analysis of public services consists of dealing with the following problems, bearing in mind the previously discussed criteria: do we have to modify the current locations? Which facilities should be eliminated? Where do we place those that need to be created? What can be gained from the redefinition of the zones served by existing facilities? It is now possible to provide answers to these questions with the help of scientific decision support systems. The purpose of this chapter is to discuss basic modeling principles of the public services location problem. The rest of the chapter is organized as follows. In section 3.2, we present the ingredients of the model. First, we propose a typology of public services that takes into account most practical situations. Secondly, we describe how to model demand and potential supply for a given service. Thirdly, we discuss some relevant criteria of optimality. Finally, we present ways of performing sensitivity analysis on the proposed solutions. Section 3.3 includes the formulation of a basic model widely used in practice, the *k*-median. Section 3.4 proposes a case study, while the conclusion suggests research paths in order to improve the methodological approach.

3.2. The modeling approach

Searching for the best spatial adequacy for a public service to a given population can be summed up in three fundamental questions: (1) *what needs must be satisfied?* This question encompasses the definition of demand and its geographical distribution over a given territory. (2) *What are the possibilities of response?* We are now on the supply site, which is defined both in terms of existing establishments and by the technically and economically feasible potential modifications. (3) *How can we adjust supply to demand?* In other words, we want to suggest solutions with their respective advantages and drawbacks for the population as well as the public agencies concerned.

These three general questions define the *structure* of a public services location study. This relies on the following steps:

(a) a general reflection on the nature of the considered service as well as its institutional and local characteristics;

(b) modeling the demand for the service in terms of intensity as well as of geographical distribution;

(c) the definition of the potential supply, which incorporates the evaluation of the current situation as well as its possible modifications;

(d) the adjustment of supply to demand in order to optimize geographical, economic, financial, environmental, etc. criteria measuring the performance of the service;

(e) an assessment of the solutions proposed by the model including a sensitivity analysis to the data introduced in the model and the consideration of elements that are hard to quantify.

3.2.1. *A typology of public services: an attempt*

The expression "public services" covers a broad range of situations and activities so different from each other that it would be useless to attempt to produce a general location model for all possible cases: the operating mode of the service determines the selection of the location model. Therefore, we propose a basic classification of public services. This typology, far from being exhaustive, is mainly based on the nature of the public service and on the characteristics of the trip between the service and the user. In operational terms, this step is essential for understanding the essence of the location problem.

Public services can be categorized as *standard* or *emergency*, depending on whether the user's request must be processed with or without delay. In the case of emergency (ambulances, firefighters, police, etc.), the time before the service arrives on the spot, including the detection of the alarm, the transmission of the call, the travel time to or from the service, is the major criterion measuring the performance of the service.

The service is *fixed* when the user travels toward the service (schools, sports centers) and *mobile* when the service travels toward the user (home medical care, mail distribution). This distinction is particularly useful to determine the terms of the trips to the service (who decides, when, where, how, who pays, etc.).

The service is said to be *demand elastic* or *inelastic*, depending on whether the number of trips made by the user depends on the transport cost and on the pricing of the service (swimming pools, socio-cultural services) or not (primary schools, emergency services). This is all relative to how the demand for service from potential users will be modeled.

Services subjected to *congestion* can be distinguished from those that are not. In the first case, the use of the service by one person makes it unavailable to another person, or at least is liable to alter its quality. Public transport or the use of road infrastructures are good examples. In the second case, the service remains completely available for any user needing it, no matter how many other persons are using the service at that moment. This is closely linked with the notion of public good in economics and can be illustrated by the production of radio programs.

The *mandatory* characteristic of some types of services has two aspects: a given category of population can be compelled to use it (e.g. primary schools) or the service can be obliged to respond to any demand it receives (weekend medical duty). The other types of services are called *optional* (for example, recreational services).

For some fixed services, the user freely *chooses* the place where she/he will consume (libraries, swimming pools), whereas for other services, the provision place is imposed (certain public administrations). In the first case, we have to make a distinction between situations where the user has the possibility to patronize various service centers and those where they can only select a single center.

Any service has an impact on the physical and human environment. We can separate *attractive* or *repulsive* services based on the nature of emitted externalities. In this way, a hospital can generally be considered as attractive, whereas a garbage station or an incinerator would be qualified as repulsive. Things can get more complicated when we consider that the people suffering the disadvantages of a situation are not necessarily the ones benefiting from it.

The "geometric" nature of the service is also crucial for the classification and hence for modeling. Some services are *punctual*, meaning that they are provided at places that, at a given scale, can be represented by a point; a hospital, a parking garage, a cultural center, a football stadium are typical examples. Other services operate on *networks* (communication routes, energy or water supply systems, etc.) or *areas* (parks, natural reserves, etc.).

Classifications made with this grid are obviously not universal and largely dependent on the institutional context in which they are considered. For example, the period of mandatory school attendance is different in each country and the choice of the school may be totally free in some countries and subjected to compulsory allocation rules in others. Let us also mention that a service can belong to several categories. A good example is the postal service, which is a fixed service (counters) as well as mobile (postman), mandatory (recorded delivery) and free (other postal transactions), attractive (small post office) or repulsive (sorting office), etc. [THO 86].

3.2.2. *Estimating demand*

After having acknowledged the type of service, we start with *locating demand*, i.e. by mapping the users by punctual observations if there are not numerous, or by spatial aggregates of individuals (street sections, statistical wards, districts, communes, etc.). The spatial aggregation level depends on the geographical scale at which the problem is studied and on data availability. As in many other spatial models, it can greatly influence the operational results (see, e.g., [CAS 87, CUR 87, FRA 04, HOD 93]).

The *intensity of demand* has also to be estimated for each demand point. This quantity depends on the nature of the service: number of visits, number of calls, number of interventions, number of clients, etc. within a given period of time. This can be done for different categories of users, for different degrees of accuracy, for different periods of the day, week or year, etc. We can disaggregate demand, for example, by demographic categories (age, sex), by social and occupational categories, by sub-categories of service, etc. In order to determine the demand, we must often proceed by *surveying* users of the service and/or concerned public or private agencies. When these surveys cannot be done, the quantity demanded by the total population or by a fraction of it (the number of school-age children for the schools, for instance) can be roughly approximated. This estimation is reliable if the studied service globally addresses the whole population or certain age group categories and the demand is inelastic. However, when the demand is elastic, it becomes advisable to build an *individual demand function*. By this we mean a relationship between the quantity of demand and its determinants such as the attributes of the service, of the user, and of what separates them geographically (basically the distance). This function may include characteristics of human behavior such as multipurpose trips or the possibility of a user visiting several provision center of the service. In this case, it is obviously important to proceed to surveys to collect the information necessary for the econometric estimation of the demand function.

Estimating the temporal variation of demand, i.e. the time distribution of calls or visits to the service, is also important for the planner. The use of a service unavoidably generates some level of congestion and the public authority must take this fact into account by scheduling an appropriate response capacity, for example, by using adequate hours of operation, quantity of facilities or the number of staff. Since the future is rarely known with certainty and the decisions must sometimes be made for the long term, it is also important to develop scenarios representing plausible evolution states for the demand for a service, in terms of quantity as well as in terms of locations.

3.2.3. *Analyzing supply*

Analyzing supply first means to *inventory* the existing facilities: their number, their exact location, their size as well as their operating mechanisms. This information is often obtained by surveying the decision-maker. Next, we must *identify* the potential locations, i.e. sites fulfilling the technical, legal, environmental, etc. conditions for possibly accommodating a new facility. Making the list of potential supply sites is more important than what would seem *a priori* because it conditions the relevance of the proposed location solutions. Among other factors, this choice must take into account the impact on the environment and the population. Some services, although globally necessary to the community, are in fact a source of nuisance affecting their immediate surroundings (visual, olfactory or sound pollution; congestion). Two approaches are then possible: either eliminate from the list of potential locations those where locating a facility would be too harmful or raise too much opposition, or evaluate as best as possible the social cost associated with the location of a facility in such a center. Last but not least, it is necessary to determine the set of requirements (called "constraints" in the sequel) that should be fulfilled by any redeployment proposition. These constraints can be of different types. Some involve the *number* of units to install: in general the agency has a limited budget that will be shared among a number of spatial units. This number has to be defined and will depend on the geographical dispersion of the users (the larger the dispersion, the more units to be deployed) and on the economies of scales due to concentration (if these are high, a smaller number of large units will be favored). Other constraints involve the facility's *capacity*: it is possible to impose a minimum capacity that avoids a solution below a minimum profitability threshold (e.g. minimum school attendance) or a maximum capacity to avoid harmful effects of congestion and/or of gigantism (e.g. maximum number of beds per day care center). In most situations, the implementation of capacities will be done dynamically starting from an initial situation and by proceeding with expansions and/or removals of capacities alongside openings/closings of sites. Finally, the spatial organization of the services is generally paired with the implementation of a *hierarchy* of sites, as in the case with health systems where very expensive equipment or very specialized medical teams are concentrated in a few areas, leaving dense territory coverage for more common situations.

3.2.4. *Adjusting supply to demand*

Once supply and demand are defined, we proceed with their mutual adjustment. This implies that we are able to evaluate the possible location policies (the current configurations as well as the redeployment projects). It is physically impossible to make comparisons by pairs since the number of possibilities can be very high. It follows that we must use an *operational location model*. This model is consists of one (or several) objective function(s) and a set of constraints. There are many types of location-allocation models and many variants (for examples, see [DAS 95, DRE 95, DRE 02, HANS 96, LAB 95]). Only the general structure of this type of model is discussed here and a simple example is proposed in section 3.3.

The *objective function* assesses the performance of any spatial configuration of facilities and translates, in a formal manner, a criterion set by the planner. In principle, this criterion should summarize the global outcome of the possible redeployment operation (the net collective advantage associated with a configuration). In practice, such an outcome includes several aspects and it is not possible to evaluate all of them. That is why the criteria that we use will settle for capturing the essential while remaining simple enough to be practically relevant.

Very often, and as we will discuss in section 3.3, the retained criterion will maximize the accessibility to the service, i.e. *minimize total travel costs between users and facilities*. For inelastic demand services, this criterion is completely relevant. Indeed, the relative advantage of a location for a user resides mainly in the cost (time) of the corresponding travel to/from this site. In the case of elastic demand services, the user benefit depends on its distance to the service and cannot be expressed only by the travel cost. Another criterion must be used. In accordance with the principles of cost-benefit analysis, we propose measuring the benefit associated with a location by the *consumer surplus*. This is defined as the difference between what the consumer is willing to pay for each unit and what they actually pay; it is represented by surface *S* in Figure 3.1. In this figure we have indicated a demand function for a certain good $p(q) = 50 - q$. Let us presume that the cost of the good is 20, which means that we can acquire 30 units. In order to obtain the first unit of the good, the consumer is willing to offer 50 but gets it for 20; therefore they save 30 on this unit. To obtain the second unit, they are ready to pay 49 and since they always pay 20, they save 29. Continuing with this thinking until the 30th unit, we see that the surface *S* graphically measures the "gain" of the consumer. This concept of surplus is fundamental in public economics. In our public services location context, we retain as criteria the maximization of the sum of the surplus benefiting users or groups of users.

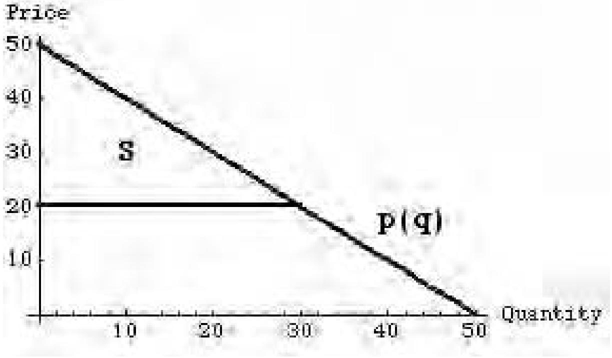


Figure 3.1. *Consumer surplus*

These criteria only consider the users. We still need to integrate the costs borne by the public agencies, such as the *construction costs* (generally expressed in terms of equivalent annual cost) and the too often neglected annual *operation costs* of the facilities. If cost evaluations are accurate, we merely need to add these costs to the criteria as previously defined, otherwise we can either integrate them to the model as constraints or, in a more radical way, use a different model where public costs are minimized subject to service accessibility constraints. As discussed in the introduction, the solutions will be the result of a *compromise* between a social accessibility goal and the investment costs needed for operating the system.

The *constraints* imposed are those explained earlier with regard to the number, size and hierarchy of facilities. They are introduced in the model in such a way that any solution respects them. Other constraints can also be added. In the case of inelastic demand, we naturally want to satisfy the total demand. We must then make sure that the different constraints are consistent between each other; otherwise the problem studied does not make sense. For example, requiring that the total demand is satisfied while limiting the number and the size of the facilities can be impossible. In this case, we must relax certain constraints, maybe by penalizing a user's non-service or by assigning another objective function to the model such as maximizing the number of service users.

The minimization of travel costs often leads to choosing locations which favors users grouped in large population centers at the expense of geographically spread out users. Therefore, there is a risk of generating a significant discrimination in service accessibility within the population considered. If such a situation is found to be intolerable, we should either integrate additional constraints to the model to enforce more equity or opt for an alternative criterion.

The first method can be very simply implemented, by requiring that each user has at least one facility within a certain threshold distance. We are again facing a bicriterion problem in which we show the compromise between an efficiency objective, expressed by the minimization of the total distance, and an equity objective, represented by the distance between the service and the worst-off user. Note that this can lead to an infeasible solution where the available budget (here the number of locations to build) is insufficient to guarantee that each user finds a unit within the prescribed distance.

As for the second method, we can consider minimizing the largest distance between a user and the service (similar to Rawls' criterion in public economics). However, this criterion presents several drawbacks. First, it can lead to very expensive solutions when there are small groups of remotely located users, as each of them will receive its own facility. Secondly, this criterion strongly depends on the definition of the demand. Let us illustrate this with a simple example about three users located at abscissa 0, 0.99 and 1 of a line segment. The optimal location, according to Rawls' criterion, is at point 0.5. Considering that the two closest points constitute in reality a geographical entity to which we attribute a weight of 2 and an abscissa of 1, the optimum is at the abscissa 2/3. Since such differences are not acceptable, we can consider a substitute to the Rawls' criteria consisting of raising all the distances to an exponent higher than 1. This will increase the weight attributed to marginal groups while taking all groups into account. Going back to the previous example, we can easily see that the problems caused by the spatial aggregation of the demand are fading. Finally, when we increase the exponent, we quickly reach a solution very close to the one obtained by Rawls' criteria. Other equity criteria can also be considered. In this way, the distribution of distances between the service and users can provide precious indications on spatial equity, for example, with the Lorenz curve.

Finally, another way to address the problem could be to minimize the investment in the system, all the while guaranteeing to each user a minimum service threshold. For example, we can look for the minimum number of fire stations so that any population center in a rural area will be at less than 20 minutes from the closest station.

3.2.5. *Evaluating the solutions*

The location model provides a redeployment plan (the solution) optimizing the settled criterion, while respecting the explicitly formulated constraints. The solution simultaneously indicates the *location*, the *number* and the *size of the facilities to be implemented*. Taking into account an existing situation, this obviously means locations to be eliminated, added or relocated. The solution also provides the allocation of users to the different facilities so that it is possible to map the optimal *service areas*. This map is particularly useful for mobile services, as these can define their own operational zones. It is similar for fixed services that impose the consumption places. In practice, the model may sometimes indicate that the current service areas are inefficient and consequently suggest a new partitioning of the territory whose implementation is most often not very costly (police, fire department, emergency medical services, etc.).

The evaluation of the solutions of a model also includes sensitivity analyses. As previously explained, the application of a model to a real-world situation relies on the estimation of several elements: costs and transport time, implementation and operation costs, selection of potential sites, evaluation of demand, etc. These estimations are always to be treated with caution, as they are dependent on taken or received measurements, the latter being often contaminated with all kinds of errors. Moreover, the context in which the service exists evolves with time in unpredictable ways. We must therefore define the elements that are the most likely to generate a modification in the solutions of the model, for example, the range of variation of these elements in which a stability of solutions is observed, the possibility of major system modifications by the migration of one solution to another, etc.

The methods mentioned above are fundamentally unicriterion, in the sense that they generally favor one performance criterion. However, evaluating the optimal solution of a model according to other criteria judged less important might be interesting. This could eventually lead to an arbitrage between alternative solutions or to the selection of suboptimal solutions.

3.2.6. *Methodological perspectives*

If the models just mentioned have proven their advantages, they are still somewhat simplistic and require further research. For example, many of these models are purely *static*. In reality, the investments have a certain durability and the geographical distribution of the demand addressed to a service is likely to vary throughout the equipment service life due to modifications of the demographic structure, migrations, people's needs, etc. Furthermore, since forecasts are generally uncertain, it would be advisable to incorporate dynamic stochastic elements within the models. Finally, new locations can also cause differences in the demand generation mechanisms.

Very simple assumptions are also made on the *behavior* of users (for example, the assignment to a single facility). It would be advisable to include more realistic, although more complex, behavioral aspects in location models. For example, the reputation of the staff may also differ according to the places and users may decide not to go to the closest facility but a more remote one that they prefer. Similarly, the user may vary in the facilities they patronize, for instance in the case of a multipurpose trip. In many situations, the units or departments do not receive the exact same equipment: every hospital does not have a scanner or a heavy surgery infrastructure. Users are confronted with a hierarchy of services. Incorporating this *hierarchical* aspect in models remains a major challenge. Finally, we should be inspired from recent developments in this sector since the problems are mainly of *multicriterion* nature [ZOL 92, DRE 02].

More attention should also be paid to *data* acquisition and processing methods. Econometric processes for demand analysis should be refined as well as behavior and travel (congestion, etc.) modeling. The *decision-making* and *financing* context of equipment should also be analyzed in greater detail.

Finally, it is important to note that a location model is nothing more than a simple *evaluation device* and in no way an *oracle*. The model can – and must – be used as an exploration tool for all possible decisions: with different estimates of supply and/or demand, with various objective functions, with and in the absence of such and such constraint, etc. Each use will lead to different solutions, which can then be compared, between each other. In this way, it is possible to assess the *sensitivity* of the results to the data, objectives and constraints occurring in a collective equipment location problem. Decisions can then be made with a better understanding of the issues due to these techniques.

3.3. A prototype location model: the *k*-median

In this section, we formally present a typical location model: the *k*-median, which is undoubtedly the most frequently used model for the location of public services. This model will be used in the example developed in the next section. Let us assume that the demand for a service arises from a finite set of points noted i ($i = 1, \ldots, I$). The demand at point i is noted w_i; it is assumed to be inelastic to supply and constant in time. An inventory of potential supply sites leads us to retain a finite set of places $j = 1, \ldots, J$. The unit transportation cost between demand point i and site j is written t_{ij} and, depending on the services, it may correspond to a distance expressed in meters or kilometers, to an economic cost, or to a travel time. We assume that the service has enough budget to open k centers and that the building and operating costs of these centers are independent of their location.

We introduce two sets of variables. First, for each value of index j, we define a variable y_j having a value of 1 if a center opens at site j and a value of 0 otherwise. Second, for each pair (i,j), we create a variable x_{ij} valued at 1 if the demand arising at i is satisfied by site j and valued at 0 otherwise.

The location criterion is the total transportation cost, which can be expressed as:

$$C(y, X) = \sum_{i=1}^{I} \sum_{j=1}^{J} w_i t_{ij} x_{ij} \qquad [3.1]$$

The *k*-median model consists of finding the location of k centers as well as the allocation of the demand points to these centers in order to minimize the total transportation cost. The model can be formalized as follows:

$$C^* = \text{Min } C(y, X) \qquad [3.2]$$

subject to:

$$\sum_{j=1}^{J} x_{ij} = 1, \quad i = 1, \ldots, I \qquad [3.3]$$

$$x_{ij} \leq y_j, \quad i = 1,...,I; \ j = 1,...,J \tag{3.4}$$

$$\sum_{j=1}^{J} y_j = k \tag{3.5}$$

$$x_{ij} = 0 \ \text{ or } \ 1, i = 1,...,I; \ j = 1,...,J \tag{3.6}$$

$$y_j = 0 \ \text{ or } \ 1, \ j = 1,...,J \tag{3.7}$$

The first group of constraints ensures that each demand point is assigned to one and only one supply site; this guarantees that the demand for the service is fully met and, incidentally, implies that the service areas do not overlap. The second group of constraints imposes that x_{ij} is necessarily equal to 0 when y_j is equal to 0. In other words, we prevent any allocation to a site in which a facility has not been previously opened. The next constraint limits the number of open sites to k. The last two groups of constraints limit the set of possible values of the decision variables.

It is proved in mathematical programming that the k-median problem is "NP-hard". In order to understand the significance and implications of this property, we need some definitions. An *exact algorithm* is a procedure that guarantees that the obtained solution is optimal. The procedures resulting in good but not necessarily optimal solutions are said to be *heuristic*. Next, it is useful, in order to evaluate the speed of execution of an algorithm, to take a measure that is independent of a particular computer and dataset. The temporal performance of an algorithm is characterized by an upper bound on the number of operations required for solving the most difficult problem. This bound is expressed in terms of the size of the problem (for example, I and J in the case of the k-median). We then distinguish between *polynomial* algorithms (whose execution speed is always proportional to a polynomial function of the size of the problem) and *non-polynomial* algorithms (whose execution speed in certain difficult cases can be proportional to an exponential function of the parameters). A combinatorial problem is said to be "NP-hard" when we can establish that the existence of an exact algorithm of polynomial resolution would imply the existence of polynomial exact algorithms for notoriously difficult problems such as the traveling salesperson, the quadratic allocation or the decidability of Turing machines. It is commonly supposed, although never formally proven, that the chances to demonstrate the existence of such an algorithm are scarce. However, this negative result is inconsequential in practice: there exist heuristics that function very well and exact algorithms which, although in theory are pretty "bad" (i.e. the execution time in the most difficult cases is exponential in I and J) can nevertheless efficiently solve good sized problems. Hanjoul and Peeters

[HAN 85] proposed such an algorithm for the *k*-median, for example. The reader can refer to Labbé *et al.* [LAB 95] for an overview of techniques used to solve location models, or to Wolsey [WOL 98] for more general integer programming problems.

3.4. An example: recycling centers

Numerous applications with this type of approach can be cited, making these models operational decision support tools in terms of regional development, local organization or land-use planning (see [RUS 88, RUS 89], for example). Most of the real-life applications have led to better spatial organizations and to significant savings. As example, we can mention a few projects from our research team in Belgium: modifications of the intervention areas of fire departments [RIC 90] or mobile permanence of the constabulary [THO 91, THO 93] to improve the coverage of a region, the rationalization of the number of ambulances with better locations [NEV 82], the harmonization of primary school locations in an urban agglomeration by taking into account the existence of several school networks [BEG 89, DEK 82], the location of post office counters in urban environments taking services hierarchy into consideration [THO 84, THO 86], optimal spatial organization of public libraries in urban environments [BEG 92a], location of daycare centers [FAT 82], swimming pools [KER 82], dentists [BEG 87, BEG 91], the location of recycling centers at local level (municipality) in order to minimize the cost of transportation as well as the pollution that the residents have to endure [FLA 98, FLA 02, LAU 97], the location of hospitals in Rwanda [QUE 04] and finally the spatio-temporal planning of a school network in a regional environment [ANT 00]. The methodological potentialities and the application possibilities are considerable. As an example, a very simple form of the *k*-median is used in the following section for planning a public service on a regional scale.

3.4.1. *The problem: the optimal location of recycling centers*

Let us consider a regional planning problem: the optimal spatial organization of recycling centers (or drop-off centers) in a Belgian province. All the ingredients necessary to the discussion of the planning process are present in this example: the representation of the geographical environment by a discrete set of points, the estimation of the demand in each representative point, the interpretation of the results in terms of locations, of optimal number of facilities, and of service areas, and, finally, the discussion of equity and efficiency notions.

The problem of household waste, of its reduction and its recycling has been very much relevant in the last few years and we do not expect this to change; it is in fact necessary to find solutions – particularly spatial solutions – for a healthy and durable control of the environment. It is thus necessary to provide the population with the possibility of actively participating in the reduction of the volume of waste. One of the means consists of installing an efficient network of recycling centers. The proximity of this type of service is required to encourage households to visit them regularly, but at the same time we must make sure to limit the impact of the nuisances that such services can generate. The geographer's expertise at this level is something to take into consideration.

The *recycling center* is a fenced area, under surveillance and specially laid out to enable people to deposit their household waste free of charge. This type of installation aims to promote recycling where, at the source, products must be sorted and put in appropriate bins. This type of service makes it possible to reduce the volume of household waste picked up at home by waste collecting and promotes recycling. Where land is available, creating a recycling center is relatively cheap to the community compared to home waste collecting [FATH 89]. The spatial organization problem of such facilities is at local as well as at regional levels. Locally, it is important to look at the possible nuisances generated by the service and therefore to consider a location that is the closest possible to the users while minimizing pollution effects. However, surveys have revealed that the nuisances generated by this type of facilities are low and the problems have mostly been about the noise (trucks emptying the containers and local increase in traffic) [FLA 98]. Regionally, the problem consists in the first place of finding the optimal number of recycling centers necessary to cover the needs of the population, to compare it with the current number and then to define the best associations of municipalities that would share a center (service areas of each site). We refer here to this last type of analyses.

The geographical environment studied here is the "Walloon Brabant"; it corresponds to a province to which, for administrative reasons, a municipality (Braine-le-Comte) was here added (see Figure 3.4). We end up with a studied area made of 28 municipalities located south of Brussels and characterized by a strong peri-urbanization process. We can reasonably assume that each resident goes to the closest recycling center and that the nuisance generated by this type of installation is local (micro-scale) and imperceptible at the studied spatial scale (meso-scale) (see, for example, [FLA 98, FLA 02]).

The use of the k-median proves to be fully justified: it makes it possible to minimize the weighted sum of distances traveled under a set of constraints. Its application presumes however: (1) the definition of the I demand points and of the demand weights w_i in each of these points; (2) the definition of the J potential location sites and the number k of facilities to locate/to build; (3) the construction of a distance matrix t_{ij} between the recycling centers and the users of the service.

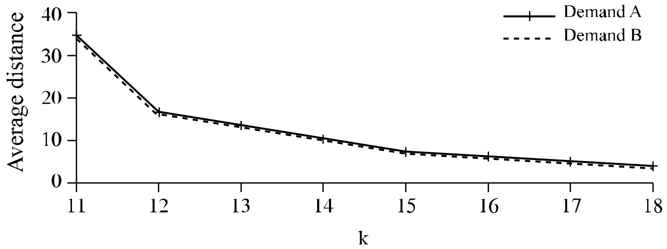
In our example, the *demand* for the service is located at the center of each of the 28 municipalities (communes), and the demand in each i is estimated either by the total number of residents (A) or by the residents owning a car (B). All the municipalities are represented by a point; each point is both a *demand* and a *potential supply site* ($I = J$). These nodes are linked together by arcs representing route segments. Each arc has a value attached to it, representing the length of the road. *Distance* is measured on a road map and weighted by the estimated average speed by road category. t_{ij} is therefore an estimation of the necessary time to travel from one point of the network to another. Special attention has been given to t_{ii}: intra-communal average distances replace these distances. Since we know that the number of recycling centers will increase in the future, we here apply the k-median for a number k of recycling centers varying from 11 (current situation) to 18; this last number corresponds to a feasible and quite realistic goal for public authorities.

3.4.2. *Results of the model*

We now calculate the optimal number of recycling centers, their locations and the service area of each site in order to maximize accessibility of the users, while respecting budget constraints as well as the specific requirements from the users and the service.

Figures 3.2 and 3.3 show the variation of k (the number of facilities) with the *average distance* traveled by the users and with the *maximum distance* separating the worst located municipality in relation to the closest recycling center for both ways of measuring demand (A, B). As expected, both quantities decrease when k increases. The average distance significantly decreases when we go from 11 to 12 sites (34% reduction) regardless the method for estimating demand (A, B). When we go from 11 to 18 sites, the average distance is reduced by 75%. If on the other hand we favor equity (Figure 3.3), the situation with 14 sites, or even with 16 sites has the higher improvement rate. Over 16 implementations, the maximum distance does not decrease anymore. The method of estimating the demand (A, B) does not affect operational results.

Figure 3.4 compares the 11 current locations (black circles) with 11 *optimal locations* (grey squares). The method for estimating demand (A, B) does not affect the results in terms of location choices. Seven current locations correspond to optimal locations. The four "extra locations" are probably due to the fact that, in our application, we plan for the organization of all the recycling centers, whereas in reality and originally, this type of public service corresponds to a local initiative and not to planning on a regional scale.



Demand A = number of residents. Demand B = number of residents with a vehicle

Figure 3.2. *Variation of the average distance traveled by the users with the number of recycling centers k, depending on the demand estimation method*

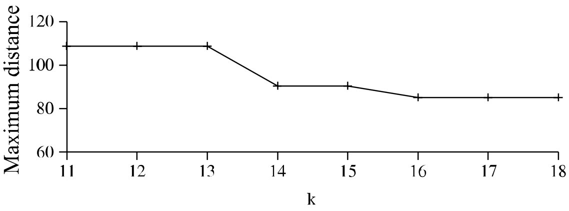


Figure 3.3. *Variation of the maximum distance traveled by the users with the number of recycling centers k (no difference depending on the demand estimation method A or B)*

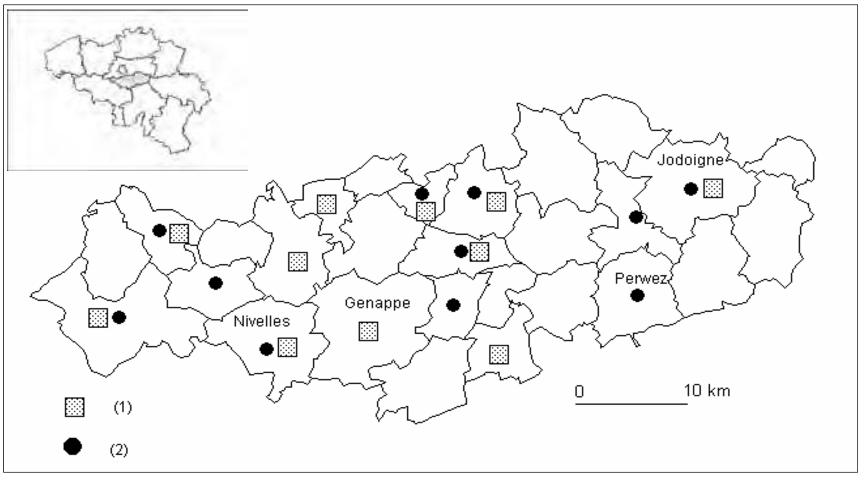


Figure 3.4. *Optimal (1) and current (2) locations of 11 recycling centers (k = 11). Top left: location of the studied area in Belgium*

The optimal allocation of each demand point *i* to the closest site *j* is another useful result of the *k*-median. Mapping this result enables us to design *service areas* and in our case, associations of communes that could be created for the use of the same recycling center. A map can be drawn for each value of *k* and this for different planning situations. We can imagine fully optimal (new) situations or optimal locations constrained by the inclusion of the existing centers. Similarly, we can try to see how a regional system behaves by removing implementations and/or by opening new sites. Finally, the additions (or removals) of centers can be done one by one or in groups. As an example, we have illustrated in Figure 3.5, the 11 optimal locations suggested by the model; the lines indicate the assignment of the communes where no recycling center was planned. Figure 3.6 illustrates the situation where $k = 14$, i.e. the 11 current sites, to which 3 new optimal sites were added. By doing so, we suggest new sites to open first in order to improve the efficiency and the equity of the current system (decrease in average distance *and* maximum distance).

Note that, as in many other case studies (see e.g. [THO 93, THO 02]), the differences obtained in operational terms do not vary much according to the estimation of the demand method (here: *A* and *B*), with the exception of a few slight differences in terms of allocation (and not of location) when *k* increases ($k = 18$, not shown here). Finally, it is important to mention that in the example presented here, and because of the significant level of spatial aggregation of demand, no maximum distance constraint for getting to the service was applied.

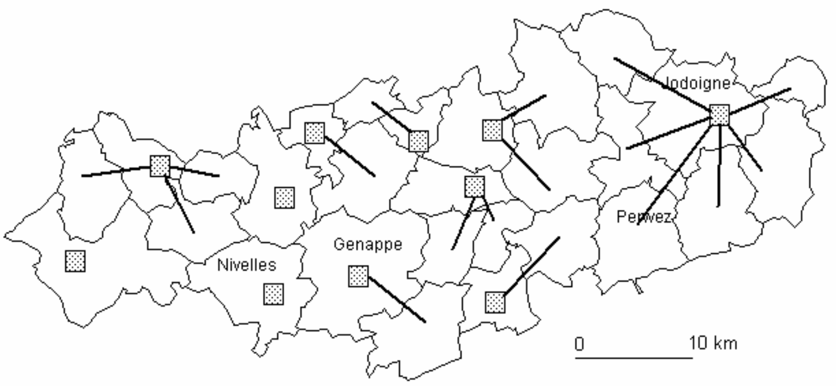


Figure 3.5. *Optimal locations and allocations when k = 11*

This example has been developed for educational purposes only; we are convinced, however, that even in its most simple formulation, this type of model is a very efficient decision support tool (see also [BEG 92b, FLA 02, THO 86, THO 04]). The *k*-median model presents the advantage of being able to incorporate multiple increasing complexities, for which the inclusion must be pondered on a case-by-case basis. It indeed often requires large data collection, or even alters the structure of the model itself that should be evaluated in relation to the results of more simple formulations that already usefully orient decisions [ANT 00, BEG 92b].

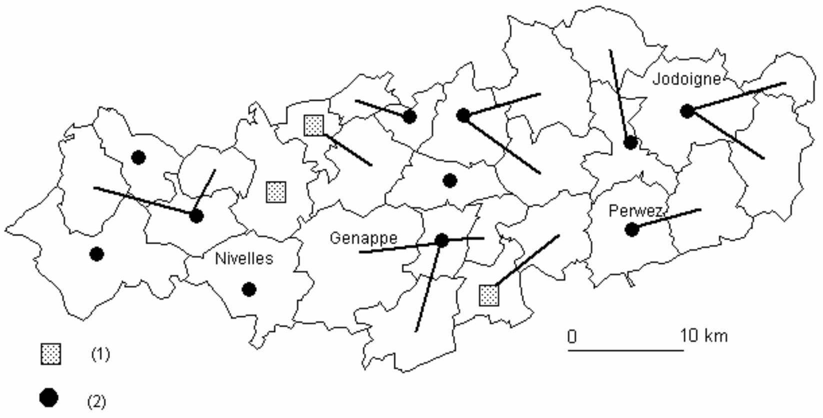


Figure 3.6. *Optimal (1) and current (2) locations of 14 recycling centers (k = 14), where the location of the 11 current sites is imposed ((2): current sites, (1): additional sites)*

3.5. Conclusion

This chapter did not have the ambition of presenting an exhaustive overview of all location-allocation models because the subject is much too vast (see, for example, [HANS 96], [DAS 95] or [DRE 02] for more detailed examination). Instead, we have attempted to lay out some directive lines and to illustrate them with the help of a simple example. A general methodology has been proposed for executing a cost effectiveness analysis taking into account constraints of different nature.

In principle, all location models emphasize the fundamental dilemma between financial cost and accessibility. In other words, they lead us to reconsider any facility spreading policy for a more balanced policy that will consider financing imperatives, while still being concerned about providing quality service to users. All areas cannot be equally equipped: the cost of such a solution is excessive so it is better to make the appropriate arbitrage by sacrificing certain sites to ensure quality of service to others. As long as everyone is free to travel and to reside in their area of choice, it does not seem reasonable to require that the community supply all the public services to everybody everywhere. Residential locations reflect, at least in part, a preference of households for certain areas endowed with specific advantages, despite their possible lack of equipment. In such a situation, the possible arbitrage can be analyzed with the help of location models since they enable us to predict the consequences of policies retained for the different stakeholders.

Along with this fundamental aspect, the location models also enable us to evaluate the social cost that a policy of geographical equity carries. Any policy focusing exclusively on the lot of the spatially worst-off citizen generally results in higher costs for everyone. This problem must be brought back to what was mentioned earlier: the legitimate desire to live in certain areas does not imply the right to the use of all public services. This comment leads us to consider the location problem from a wider angle by not restricting ourselves to heavy equipment, but by offering alternative tools as a surrogate for users with low accessibility. In this perspective, the presence of more competitive and more active housing markets will bring a reduction of disparities in the treatment of users. In fact, if, as we can expect, the rents reflect the larger proximity to public services and the differences in accessibility are partially offset by the opposite differences in rents. Landed capitalization will help in the design of an (partial) overhaul of local taxation systems that should rely more on property tax than is currently the case, in order to ensure efficient financing of local purpose public services (see [DUR 96] for more developments).

From an operational standpoint, the multiplicity of models makes it possible to test the strength of the results provided by each model by comparing them, or even combining them. The recent development of very quick algorithms enables us to process a very large problem and to do a comparative study of alternative scenarios. The location models constitute excellent decision support tools. Their goal is not to replace the deciders, but, as we have mentioned, to enhance their analysis field. As always, their usefulness depends on how we use them. They are an imperative part of the geographer's "toolbox".

3.6. Bibliography

[ANT 00] ANTUNES A., PEETERS D., "A Dynamic Optimization Model for School Network Planning", *Socio-economic Planning Sciences*, 34, p. 101-120, 2000.

[BEG 87] BEGUIN H., IPANGA T., "La localisation des dentistes dans un milieu rural", *L'Espace Géographique*, 16, p. 259-264, 1987.

[BEG 89] BEGUIN H., DECONNINCK J., PEETERS D., "Optimiser la localisation des écoles primaires: le cas de Mouscron, Belgium", *Revue d'Economie Régionale et Urbaine*, 5, p. 795-806, 1989.

[BEG 91] BEGUIN H., IPANGA T., “Private versus Public Service Location Behaviour: the case of rural dentists”, *Annals of Regional Science*, 25, p. 115-130, 1991.

[BEG 92a] BEGUIN H., THOMAS I., VANDENBUSSCHE D., “Weight Variations within a Set of Demand Points and Location Allocation Issues: a case study of public libraries”, *Environment and Planning A*, 24, p. 1769-1779, 1992.

[BEG 92b] BEGUIN H., “La localisation d’activités banales”, in *Encyclopédie de la géographie*, Bailly A., Ferras R., Pumain D. (ed.), Economica, Paris, p. 515-531, 1992.

[CAS 87] CASILLAS P., “The Aggregation Problem in the p-median Model”, in *Spatial Analysis and Location-Allocation Models*, Ghosh A., Rushton G. (eds.), Van Nostrand, New York, p. 327-344, 1987.

[CUR 87] CURRENT J., SHILLING D., “Elimination of Source A and B Errors in p-Median Location Problems”, *Geographical Analysis*, 19, p. 95-110, 1987.

[DAS 95] DASKIN M., *Network and Discrete Location, Models, Algorithms and Applications*, Wiley Interscience, New York, 1995.

[DEK 82] DEKEERSMAECKER M.-L., THOMAS I., “Un modèle de localisation des écoles primaires en milieu urbain: l’exemple d’Etterbeek (Belgique)”, *Recherches Economiques de Louvain*, 48(3-4), p. 283-300, 1982.

[DRE 95] DREZNER Z., *Facility Location, A Survey of Applications and Methods*, Springer, New York, 1995.

[DRE 02] DREZNER Z., HAMACHER H., *Facility Location: Applications and Theory*, Springer, Berlin, 2002.

[DUR 96] DURANTON G., THISSE J.-F., “La politique foncière dans une économie spatiale”, *Revue Economique*, 47, p. 197-231, 1996.

[FAT 82] FATFAT-VAN BOXEL A., PEETERS D., “Quelques propositions pour l’implantation d’un service communal de garde d’enfants de 0 à 3 ans à Sambreville (Belgique)”, *Recherches Economiques de Louvain*, 48(3-4), p. 265-281, 1982.

[FATH 89] FATH-CHAVANNE S., “L’élimination des déchets ménagers: protéger l’environnement au moindre coût”, *Problèmes Economiques*, 2141, p. 17-23, 1989.

[FLA 98] FLAHAUT B., LAURENT M., THOMAS I., “Localisation optimale d’un service semi-polluant: le cas d’un parc à conteneurs dans une commune wallonne”, *Acta Geographica Lovaniensia*, 37, p. 397-415, 1998.

[FLA 02] FLAHAUT B., LAURENT M., THOMAS I., “Locating a Community Recycling Center within a Residential Area: A Belgian Case Study”, *The Professional Geographer* 54(1), 67-82. 2002.

[FRA 04] FRANCIS R., LOWE T., TAMIR A., EMIR-FARINAS H. “A framework for demand point and solution space aggregation analysis for location models”, *European Journal of Operational Research, 159:3*, p. 574-585, 2004.

[HAN 85] HANJOUL P., PEETERS D., “A Comparison of Two Dual-based Procedures for Solving the p-Median Problem”, *European Journal of Operations Research*, 20, p. 387-396, 1985.

[HANS 96] HANSEN P., LABBÉ M., PEETERS D., THISSE J.-F., "Facility Location Analysis", in *Encyclopedia of Economics. Regional and Urban Economics*, Arnott R. (ed.), Harwood Academic Publishing, Amsterdam, p. 441-507, 1996.

[HOG 93] HODGSON J., NEUMAN S., "A GIS Approach to Eliminating Source Aggregation Error in p-median Models", *Location Science*, 1, p. 155-170, 1993.

[KER 82] DE KERCHOVE A.-M., "Analyse de l'offre et de la demande d'un service récréatif: les piscines de l'arrondissement de Mons", *Recherches Economiques de Louvain*, 48(3-4), p. 301-336, 1982.

[LAB 95] LABBÉ M., PEETERS D., THISSE J.-F., "Location on Networks", in *Handbook of Operations Research and Management Science: Networks*, Ball M., Magnanti T., Monma C. and Nemhauser G. (eds.), North-Holland, Amsterdam, p. 551-624, 1995.

[LAU 97] LAURENT M., PEETERS D., THOMAS I., "Spatial Externalities and Location Results. A Theoretical Case Study Based on the SPLP Model", *Studies in Locational Analysis*, 11, p. 193-210, 1997.

[NEV 82] NEVEN D., "La localisation des services d'urgence dans la province de Luxembourg", *Recherches Economiques de Louvain*, 48(3-4), p. 247-264, 1982.

[QUE 04] QUERRIAU X., KISSYAR M., PEETERS D., THOMAS I. Localisation optimale d'unités de soin au sein d'un pays en voie de développement : analyse de sensibilité. *Cybergéo*, article 281, 2004.

[RAW 71] RAWLS J., *A Theory of Justice*, Cambridge University Press, Cambridge (MA), 1971.

[RIC 90] RICHARD D., BEGUIN H., PEETERS D., "Fire Stations Location in a Rural Environment", *Environment and Planning A*, 22, p. 39-52, 1990.

[RUS 88] RUSHTON G., "Location Theory, Location-Allocation Models and Service Development Planning in the Third World", *Economic Geography*, 64, p. 97-120, 1988.

[RUS 89] RUSHTON G., "Applications of Location Models", *Annals of Operations Research*, 18, p. 25-42, 1989.

[THO 84] THOMAS I., "Toward the Simplification of the Location Models for Public Facilities: the Example of the Postal Service", *Papers of the Regional Science Association*, 55, p. 47-58, 1984.

[THO 86] THOMAS I., *La localisation optimale des services publics. Une méthode opérationnelle et son application au service postal*, Revue des Thèses, no. 11, Louvain-la-Neuve, Cabay, 1986.

[THO 91] THOMAS I., "Science régionale et aide à la décision en matière de gestion policière", *Revue d'Economie Régionale et Urbaine*, 1, p. 7-21, 1991.

[THO 93] THOMAS I., "Sensibilité du découpage spatial optimal des services d'urgence à la définition de la demande. Quelques réflexions sur un exemple mésogéographique", *L'Espace Géographique*, 4, p. 318-332, 1993.

[THO 02] THOMAS I., *Transportation Network and the Optimal Location of Human Activities. A numerical geography approach,* Cheltenham, Edward Elgar Publishing Ltd., 2002.

[WOL 98] WOLSEY L., *Integer Programming*, Wiley, New York, 1998.

[ZOL 92] ZOLLER H., BEGUIN H., *Aide à la décision. L'évaluation des projets d'aménagement*, Economica, Paris, 1992.

Chapter 4

Time-geography: Individuals in Time and Space

4.1. Introduction: why integrate "time" when we analyze space?

4.1.1. *The study of spatio-temporal processes*

Geography uses data, methods and tools which, used together, can explain the terms and rules of the organization and production of space. To bring to light underlying structures and forms of spatial organization, it is necessary to take into account the process that created these structures; that is why the "space-time" question is very often at the core, implicitly or explicitly, of geographical arguments.

Two main approaches integrating time can be followed. The first approach, which is centered on the research of models of organization and evolution of spatial entities and systems, takes time into account in order to compare and measure the evolution of the spatial structures between different periods. The second, which focuses on the time of individuals, considers space as a variable describing the successive locations of the individual's life path.

In order to illustrate the first category of approach, two examples can be given: urban models and diffusion models.

Urban growth is often studied with the help of models *implicitly* integrating time; that is the case of the economic models of Alonso [ALO 64] about underlying

---

Chapter written by Sonia CHARDONNEL.

dynamics of concentric form (inspired by the theory of Von Thünen), where a relation is built between the rent of the land and the distance to downtown. This is also the case with urban ecology models [GRA 79] (concentric, sectorial and polynuclear) where the extension of cities is explained by a competition for the space between the different social groups. There are also models where time is expressed in an *explicit* way: Forrester's [FOR 69] dynamic model is the most well-known example but, more recently, experiments using models with cellular automata have also made it possible to simulate urban dynamics (see the example of the city of Ottawa in [LAN 97]).

Similarly, spatial diffusion of an innovation [HAG 53] or of an epidemic [GOU 92] can be simulated with dynamic probabilistic models, showing with which process a technological invention or an illness will spread in time, on larger territories (see Chapter 5).

The second category can be illustrated by several studies, especially from demographers that have developed methodologies characterizing the forms of mobility:

– spatial mobilities are traditionally studied with the help of "origin-destination" matrices, where two points are known in space and time, making the measure and characterization of flows between these two points possible (see Chapter 1);

– social mobilities are studied using collection and processing methodologies of biographic data [BAU 92, LEL 99]; it is generally possible to characterize life paths and cycles, or focus on types of trajectories – professional, family, etc.

Finally, other studies focus completely on the time of individuals by describing their "time-budget". The "time-use" survey from the INSEE [DUM 00], for example, measures the amount of time people spend doing various activities of everyday life. We then know what are the predominant activities in the time schedules that structure daily rhythms. It is also possible to compare the evolution of these rhythms by crossing the different survey campaigns. Finally, the specificity of certain groups' schedules can arise such as those of women whose features are often mentioned in gender studies.

4.1.2. *For a time-integrated geography*

In this chapter, we will focus on a specific approach developed in the early 1970s by Torsten Hägerstrand [HAG 70a] who conducted the so-called "Lund school" (Sweden). Called *time-geography*, this approach presents the originality to address the geographical questions of individual actions through time in the spatial context of their environment. The main goal is to obtain a strong image of the geographical realities, by showing how the possibilities and constraints of the environment's structure influence the development and the construction of individual trajectories. Although inspired by time-budget demographic approaches, time-geography highlights the spatial dimension of the trajectories, thus joining the concerns of spatial analysis.

The diffusion of the works from the Lund school has been prevalent in Anglo-Saxon schools that have maintained close scientific relations with Swedish geography since the 1960s[1], when a renewal of ideas and methods (studies on spatial economy, diffusion simulation methods) created a rising trend, that of new geography. It is natural then that 10 years later time-geography emerges in the Anglo-Saxon geography world [CAR 78a, CAR 78b, CAR 82, THR 77] and the sociology world [GID 84, PAR 80], whereas it had very little impact on French geography [HAG 81]. Today, however, we do see a renewed interest, including from French authors, for this geography that echoes the increasingly numerous concerns of geographers preoccupied by the study of spatial practices, whether they are understood in terms of component and result of spatial systems [VAN 99] or in their social and cultural dimension [CLA 99, DIM 99].

4.2. The foundations of time-geography

4.2.1. *The premises*

Though time-geography development does not stem from the work of only one person, we will still associate its origin with Torsten Hägerstrand. The scientific sensitivity at the base of the main principles of time-geography goes back to his first personal studies. Two studies in particular have been important in defining a structure of thought which aims to consolidate temporal perspective in the field of geographical problems. This leads us to assume a longitudinal and contextual approach in both studies.

In the 1940s, Hägerstrand did a study on the consequences of the big wave of migration to the USA in a village called Asby, located in the southern region of Östergötland, South-West of Stockholm. Asby was particularly affected by the

---

[1] According to Claval [CLA 84] this connection dates back to the Stockholm international congress, in 1960, during which a symposium on urban geography took place in Lund (*Proceedings of the IGU Symposium in Urban Geography*, Lund, 1960, C.W.K. Gleerup, Lund, 1962).

migration[2] phenomena. The main questions were: what motivated the migrants (what situation were they leaving?) and how did those who stayed deal with the population loss? This problem has led Hägerstrand to concentrate especially on the *individual* as an indivisible being playing a succession of roles in society over time, according to his economic and social position and to his interactions with other individuals with whom he must communicate in order to live his life.

The second study in which we can trace the origin of time-geography focuses on the diffusion of innovations (see Chapter 5). The goal is to show how a technical innovation (the case studied concerns the agricultural technical innovations in the South of Östergötland) is spatially diffused through time. Hägerstrand built a simulation of this diffusion, based on the premise that the capacity of integration of an innovation varies from one individual to another according to his life *context*.

The common denominator between these two studies is to follow the people and the natural or artificial elements throughout their movements in space through time. This dynamic vision is centered on the individual and the spatio-temporal context in which he evolves: it has built the bases on which time-geography was developed.

4.2.2. *A certain vision of the world*

Before presenting its specific aspects as a modeling process (see section 4.3), we should insist on the fact that time-geography is first and foremost an epistemological positioning that proposes a frame of thought. This frame of thought is based on the idea that any construction of knowledge comes from a dialectical view of scientific knowledge and of the synoptic and contextual analysis of concrete practices of individuals.

To be more precise, the understanding of geographic problems is based on a certain "vision of the world": a world on the move where a set of elements and actors must compose, i.e. meet, unite or help each other in order to achieve their destiny. In this perspective, the idea is to never lose sight of the *situations* in which the study objects are grounded. *Space* and *time* are then considered as two fundamental resources, although finite, that must be distributed between the elements and actors in a given situation, with competition problems arising. This frame of thought has often been explained with a reference to the concept of the three worlds of Popper and Eccles [HAG 85]. World 1 is made up of the physical expression of the elements, i.e.

---

[2] Between 1880 and 1930, Sweden was affected by a large wave of migration to the USA. Nearly 950,000 Swedes left their country and the difficult conditions (poverty and rural overpopulation, famines) to settle permanently in the American territory. Hopes of religious freedom also explain the many exiles to the USA.

living organisms (a tree, the body of a man, etc.), inorganic elements (a rock, etc.) as well as artificial objects created by man (a chair, a car, etc.). World 2, which is harder to grasp, refers to the reality of our thoughts, emotions, perceptions, including our subjective knowledge. Finally, world 3 includes the existence of a culture made up of rules and laws established by man, of artistic and scientific references shared by a large number of people (with the same culture).

The advantage of this division is to show, in a synoptic and structured way, the complexity of the world, while insisting on the close connections, or even the interweaving, that all parties experience between each other. A man whose intention (world 2) is to go on a car trip must adapt to the physical objects at his disposal (world 1) – a vehicle, a transport network, fuel, etc. – and must, during his trip, respect a certain number of rules established by society (world 3), such as the traffic laws.

One of the aims of time-geography is to highlight the relations between the physical realities of our environment and our capacity for action. Even if we are not required to submit to the idea that world 1 strictly determines worlds 2 and 3, we must admit that it imposes limits leading to consequences on our capacity for action and thought. "We must invent a language which helps us to keep existents and events in the three Worlds together under a unifying perspective, and it is equally essential to pay attention to the bridges and the traffic which crosses them" [HAG 85].

The aim of the methodology is then to reveal the reality of these constraints coming from the "corporeality" of the human being (his lifetime is limited, so his time is rare) as well as from the limited "packing" capacity of space (two objects cannot occupy the same place in space at the same time). These require the sequencing of human activities in a specific order in space and in time. From there, the developed model revolves around three principles:

– having a holistic approach that does not deny the complexity and that brings to light the contexts;

– developing a longitudinal analysis centered on people in order to track what regulates their sequential use of space and time;

– formalizing a mode of graphical representation that leads to a reasoning that is more synthetic than analytical.

4.3. The conceptual framework of time-geography

4.3.1. *The creation of a "notation system"*

Since the goal of time-geography is to highlight the actions and processes forming the complexity of the world, it becomes fundamental to describe in a realistic way the elements involved in the development of an action: any process is localized, located in time and requires space for its realization. Time and space are considered as a unit, space-time, a vital resource for any action. This approach to reality is deliberately dynamic since it uses time and space as two elements of the same nature and dimension, as in natural and physical sciences [HAG 70a]. In order to work this conceptual formalization, a graphical "notation system" is suggested, surpassing the purely static aspect of the topographic map: "We need to rise up from the flat map with its static patterns and think in terms of a world on the move [...]" [HAG 82].

The "notation system" proposed comes from a heuristic approach, by the "visualization" of spatio-temporal realities of the studied processes. The creativity resides in the nature of the description which is neither verbal nor mathematical or even numerical. It is much more inspired by the notation system of chemical formulae and by musical writing. Indeed, as in a chemical formula, the notation must be able to show the links between elements. And just as a musical score – a staff and notes – corresponds to an acoustic sequence that flows in time, the notation must show spatial configurations as well as temporal successions [HAG 70b].

The main goal here is to give a physical reality to events that we cannot observe simultaneously or completely. A form of communication with its graphical codes and its semantic significations is then at the basis of time-geography representations. The initial principle is to consider that we are explaining a process in a three-dimensional diagram: "If we imagine the geographical space brought down to a plan, then the scale and direction of time can be displayed along a vertical axis. In this way, the movement is transformed in geometrical form. Even what is immobile in space has an expanse in time. An isolated action becomes a point with a position in space and time" [HAG 70a].

Even if the invitation to represent the events in concrete terms is tempting, we must specify that this "notation system" is first and foremost a support for reflections. We must see it as a learning tool that helps to visualize the reality of actions in their spatial and temporal dimensions, and to understand their physical position in their environment (context-situation) and continuously follow how they progress (biography). This "notation system" is highly pervaded with the physicalist thought in which the world can be described by concepts inspired from natural sciences (in the Greek sense *phusikê*) and that everything is an object. Since it is not actually possible to draw in concrete terms all the geographical phenomena studied in the form of a three-dimensional diagram, this form of notation must be kept in mind and enable us to explain, mentally at least, the reality in terms of situation and process. The "notation system" proposed is more of an illustration of the frame of thought and does not aim (at least not at first) to be an analytical tool in itself [HAG 70b].

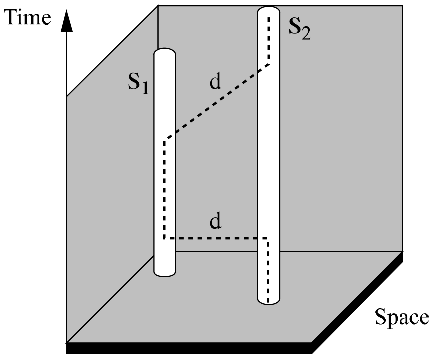
In conclusion, time-geography can be defined as a "form of language" trying to overcome the limits of mere words that are unable to express complexity, interdependences and overlapping [OBE 89]. The semantics of this form of graphical expression makes it possible to give physical visibility to potentially observable, but yet unperceived by the observer, processes and relations. The journey of an individual in an urban space throughout one day does not leave a continuous, material and observable trace. Because of this, it is difficult to grasp the individual's course as a whole in order to understand the logical relations between the sequence of the places visited and the activities done in the structure of the environment. With the proposed "notation system", we will be able to simultaneously visualize the whole journey by giving it the form of a dotted line with a reference to a temporal scale and a spatial data representation.

4.3.2. *Tools to decrypt daily life*

Hägerstrand has formalized a set of notions based on the "notation system" and has built a three-dimensional block diagram capturing the geographies of everyday life.

4.3.2.1. *Trajectory, station, project: basic concepts*

The first notion is *trajectory* (see Figure 4.1). It refers to all the concrete positions and movements that an individual will execute in space-time. The start of the trajectory of an individual can be considered as the *point of birth* and ending at the *point of death*. The trajectory of the individual can be processed on many levels: "when we focus on short length actions, lasting one or several days, we can qualify the trajectory as *daily trajectory* of the person. The daily trajectory can become weekly, monthly and annual. When we are following a person over a longer period, we can even talk about *lifetime trajectory*" [HAG 70a].



Space represents a territory, large or small, and time is represented on the ordinate axis. An individual follows a trajectory which is made up of stays in different stations *S1*, *S2* and of movements *d* between the stations.

Figure 4.1. *Trajectories and places (from [HAG 70a])*

Using the concept of trajectory of the individual means that we consider the individual as a physical unit moving in space-time. The individual is not defined by a role (consumer, driver, etc.), but in relation to the activities that he performs and to the places where he goes and the other people that he meets. A population then becomes a bundle of individual trajectories interweaved through space-time, visually forming a type of weaving pattern: “the use of the weaving metaphor for the movements of individual trajectories in space-time is not just because the weaving is two-dimensional and space-time is four dimensional. The merit of this metaphor is that the weaving image illustrates the daily routine of trajectories which tend to repeat themselves day after day, as the meeting of a trajectory with certain other trajectories also normally repeats itself as in certain weaving patterns” [AQU 92].

The trajectories of the individual are not randomly formed but are scheduled according to the structure of the environment and of the actions planned. The first factor that schedules the trajectories is made up of *static positions* where, for a certain period of time, an activity takes place. They are called *stations* and they often correspond to infrastructures that are physically stable in time such as a school, a store or even a sports court. The notion of station has a flexible scale in time and in space, just like the notion of trajectory of the individual. What can be called a station in a life trajectory perspective (for example, the town of residence) will be divided into a set of stations in a daily trajectory perspective (dwelling, stores, workplace, etc.) [HAG 70a].

The second factor corresponds to the *projects* that are motivating the action. In fact, a trajectory can be broken down in all the activities that an individual or a professional organization (a company, an administration) must execute in a certain sequence in order to reach a goal; if the individuals do not always formulate their intentions in terms of projects, the projects will nevertheless often guide the action. In practical terms, a project can be defined in short or long term: it can be for a meal preparation, the production of a manufactured product, or the education of a child. This notion is interesting insofar as we can identify the manner in which the projects are developed, how they get structured in a sequence of precise activities and especially in what measure the projects require time and space, sometimes competing for the same resources.

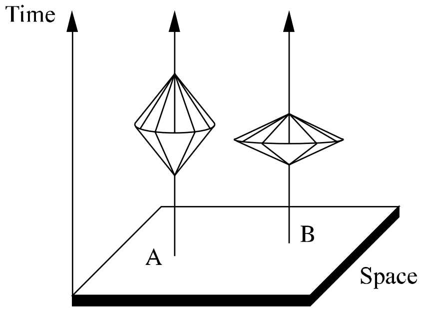
The mere graphical visualization of the trajectories of an individual paired with the notion of project can be interesting in a behavior study. Kjellman [KJE 95, KJE 03], for example, has used this description in a study on reintegration of drug addicts in rehabilitation. The goal of the study is to show how drug addicts' daily life can be reconstructed during an original and innovative detoxification program in a Swedish center. The approach consisted of following the patients in the different phases of the program by constituting spatio-temporal trajectories before, during and after treatment. The author relies on the visualization of patients' daily paths to test the hypothesis that managing spatial and temporal resources is a control factor of daily life. When the patient's main project sends him back to a socially abnormal practice (drug use in this case), the journey is completely disintegrated in relation to the environment. By prohibiting the main project and by limiting movement possibilities, the treatment established a type of minimalist and extremely routine trajectory, thus disrupting completely with the previous unstructured journeys. The "reappropriation", through routine, of daily spatial and temporal structures, made it possible to progressively reintroduce other projects (work) that would integrate into a well-ordered trajectory, in line with the social as well as spatial structures of the environment.

4.3.2.2. *Different types of constraints*

Visualization and longitudinal analysis of trajectories motivated by projects quickly demonstrate the reality of the constraints mentioned previously (see section 4.2.2), linked to the corporeality of the human being and to the nature of his relations with time and space: "in reality we come up against a lot of barriers. The individual gives the impression of being an observer caged in a labyrinth limiting the real possibilities of choice" [HAG 70a].

Identifying these obstacles is interesting because they shape the trajectories of individuals in space-time and can be explanatory elements of certain types of behaviors. Their modeling makes it possible to anticipate the spectrum of possible actions for a type of person in a given situation. Even if it is impossible to offer a comprehensive taxonomy of constraints seen as time-space phenomena, Hägerstrand defines what he calls three large "aggregations of constraints": *capability* constraints, *coupling* constraints and *authority* constraints.

Capability constraints cover everything that limits the activities of the individual such as his biological attributes and the tools at his disposal. The constraints related to the biological constitution are, for example, the fact that we must sleep for a certain length of time each day and that we must reserve a few hours for meals. These two aspects come back regularly and strongly constrain our usage of time. Other capacity constraints deal more with our limited capacity for action on the environment above a certain distance. The necessity, for example, to be settled in what we may call a *point of retreat*, often the dwelling (where we can rest, eat, etc.), imposes on individuals a short amount of time during which they can get away from this point. The distance they can reach from this point is also related to the performance of the travel mode.. The individual possesses an expanse in space and time that we call *daily prism* (see Figure 4.2): "we can say that in our everyday life, the individual is linked to an island with a specific radius… In the time-geography approach, the island that the individual can cover takes the form of a prism" [HAG 70a].



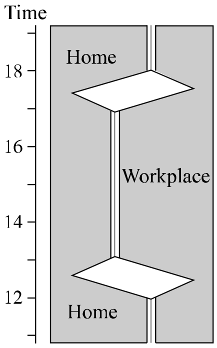
The expanse in space-time that the individual possesses during a day is illustrated here for two individuals residing respectively in A and B. These points are the bases that the individual should not leave before a certain time and to which he must return before a certain time. The external positions of the return points are determined by the maximal capacity of displacement. The points inside the prism are the only ones that can be visited.

Figure 4.2. *The daily prism*

The actual expanse of an individual throughout a day is much smaller than the potential expanse determined by his traveling capabilities. In fact the daily prism must be considered in relation to the activities that schedule the days. Work, for example, requires presence in a determined place for several hours a day. Commuting time is an important factor in the actual prism of an individual. If the workplace is at the maximum distance that the individual can reach taking into account the available time to travel, then no other place can be visited before or after work. In this case, a person with a slow transportation mode has fewer choices of workplaces than a person that can travel quickly. When the workplace is not at the maximum distance, then the day can be divided in several prisms as illustrated in Figure 4.3.

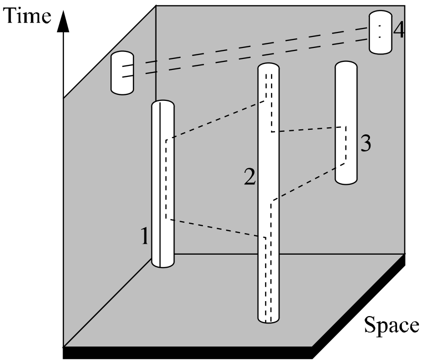
The *coupling constraints* result from coordination requirements, on the one hand between individuals and on the other hand, between individuals, tools and equipment. Most of projects (production, consumption and social interaction) require that the individuals, tools and equipment are bound together at given places at given times in what we call a *bundle of activity* or a *system of activity* (see Figure 4.4).

Ellegård focuses on notions of projects and coupling constraints in a study on the organization of the labor market by investigating, at several levels, starting from the general productive logic all the way to the reality of individual work.



The individual considered here cannot leave his home before 12:00 PM. He must be at his workplace between 1:00 PM and 5:00 PM and must be back at home at 6:00 PM. Since the travel time between his home and his work is less than 1 hour, he then has a prism between 12:00 PM and 1:00 PM during which he can do activities located within the space defined by the prism (the spatial limit of the prism correspond to the farthest places that the individual can reach, considering the time needed to return to the workplace on time). The same situation happens between 5:00 PM and 6:00 PM.

Figure 4.3. *Example of a prism (according to [LEN 78])*



Different activity bundles are represented by 1, 2, 3 and 4. Bundle 4 represents a temporal but not spatial coupling; this can be the representation of a telephone conversation.

Figure 4.4. *The web of individual trajectories into activity bundles (according to [HAG 70a])*

The application took place in the organization of automobile production, in particular in the Volvo plant in Uddevalla [ELL 91, ELL 96]. Ellegård observed the flows inside the plant and compared the efficiency of a traditional assembly line with that of a team organization where a group of individuals is in charge of the whole assembly process of automobiles. This study has led to the deployment of a "reflexive" production where the workers, in teams of two or four people, assemble cars in a fixed station. The emphasis of the organization of the entities trajectories involved in the production process (workers, machine tools, spare parts, automobile) has enabled us to argue for a conjunction constraints optimization and to prove that a holistic approach to the work of production workers was more profitable, in terms of manufacturing time, than a traditional chain assembly organization. Furthermore, a consideration on the notion of project shows an increased involvement from workers in all the manufacturing-related tasks, thus giving them more responsibility and control and, more importantly, personal pride in their work.

The third type of constraint, *authority constraint*, refers to limitation and control of access. They reveal the exercise of power in space and time. "Each inhabited territory has been divided over time in a network of small and large units, some with tight boundaries and others more vague, in which certain people of designated groups have certain rights and are authorized to make decisions and exercise control" [HAG 70a].

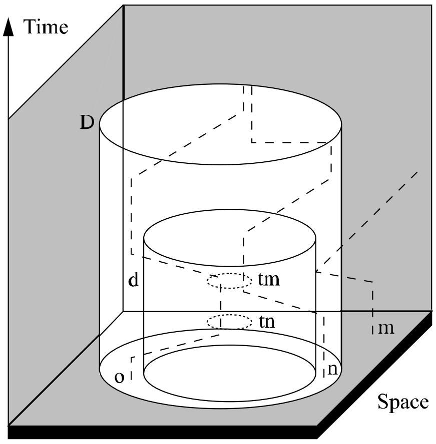
The term *domain* is used to qualify these spatial entities that are based on different grids, but that are all under the control of a relevant authority. The organization of space into domains makes sense, insofar as a certain order is necessary so that the accumulation and coexistence of individuals in a same place does not systematically lead to resource sharing conflicts. That is why our ability to visit a domain is often determined by, for example:

– the possession of a ticket (where the sale is regulated by the domain's proper authority and the validity is often limited in time) as in the case of a movie theater, a museum, etc.;

– a permanent access over a period of time, such as an identity card certifying citizenship, authorizing the presence on a national territory.

Authority constraints do not only control the physical expression of the domain, but also its administrative, political or even social reality. They account for the rules controlling space and time. In fact, the individual projects must be scheduled in time according to available time slots for different types of activities. Work hours are often the ones that will structure daily programs first, but many other domains also limit visits to certain hours of the day: that is the case with administrative or commercial services. These time slots are the result of higher decisions that apply to all the organizations under a same domain.

The individuals must then organize their activity program by navigating in a network of hierarchically organized domains. Thus, certain domains are directing the operation modes of lower domains (see Figure 4.5).



D represents a higher domain (for example, a town); *d*, a lower domain (for example, a town service or a store). Domain *d* can be regulated by *D* as long as, for example, people living in town *D* (*n* and *o* but not *m*) are the only ones who have access to the services proposed by *d*. The *tn-tm* period corresponds to the opening time hours of the service to the public.

Figure 4.5. *Domain hierarchy limiting individual trajectories*

Since the end of the 1990s an original illustration of the way in which different elements that organize spatio-temporal constraints are taken into account by public policies can be observed. In several European countries (Italy, the Netherlands, Germany and France [BOU 02]), quite a few national or local agencies for territorial planning put forward "temporal policies". They are based on actions on rhythms of life in cities and their surroundings (opening hours of services, alternate flows, etc.), so as to adapt them to the needs and timetables of the citizen (working time, family time, free time, etc.). Examples of such actions are quite varied; they concern many different aspects, such as organization of childcare at long and changing hours, the opening in decentralized public offices of only one window for different services at the same time and in the same place. Research is led on the different uses of space according to the time of day or the moment in the week, so as to enable an urban reorganization adapted to the rhythms that have been observed.

4.3.2.3. *A transversal analysis of the "three worlds"*

When we observe concrete situations, we can quickly see that the projects of individuals are not only confronted with one constraint or the other, but they must also deal with combinations of restrictions that are spread over the structure of the environment. The different constraints are therefore often intermingled and can even reinforce each other. In fact, the actions entailed by a type of constraint can sometimes bring about another type of constraint, thus putting a *shadow effect* on the future activities program. Mårtensson [MAR 79] illustrates how the decision of a change of organization of a domain, the move of the forestry engineering school in Stockholm to another town, has numerous repercussions on the organization of lower domains like the homes of the people employed at the school. The author reveals different types of consequences: the effects on work itself within the school in terms of time needed to go on a business trip to Stockholm or to other Scandinavian cities, the consequences for the other members of the family with regard to employment opportunities in the new town, the effects on the capacity to maintain a certain lifestyle in terms of recreation and of social networking still in place in Stockholm. The aim of this study was to show how individual biographies are developed in relation to their environment. The considerations of this study took place at the heart of the debate that had been going on since the 1950s in Sweden about the quality of life and social planning concerns. The author's initial premise was the belief that in such a context, the notion of "quality of life" does not only refer to the existence and permanence of high standard equipment, but also to the contents of the biographies. In order to be acceptable, the notion of quality of life must include components such as reason to live, having a social life, having a position which makes it possible to influence our own life, to make it stimulating and to use our imagination, and feeling secure. We then understand that such an aim cannot be limited to a systematic analysis of the physical structures of the environment and of the constraints that follow, but must always take into account and show the relations between the "three worlds". The aim is to formalize this interaction idea between what comes from physical reality and what comes from the mental and cultural: we need to materialize these arrangements of resources (people, equipment, knowledge, etc.) that are scheduled and controlled in order to let certain type of activity happen in given places at given times without being jeopardized by the accomplishment of another activity [HAG 85].

4.4. Time-geography in practice

The theoretical and conceptual development is combined with experimentations in several geographical spaces and areas. The different studies, which have been deliberately embedded in time-geography for 30 years, have not favored the same aspects of the frame of thought. The two following examples have been chosen to demonstrate the evolution of the approach and of the diversity of the application fields.

4.4.1. *Simulation of individual activity programs: public transport possibilities in the city of Karlstad – an application by Bo Lenntorp*

The first application presented here was done by Bo Lenntorp in 1975 and took place in the context of the first development of time-geography conducted within Hägerstrand's research group at Lund University. The aim of this application was to implement the theoretical frame of the approach by especially developing the notion of *prism*. The focus was on the analysis of daily activity programs of individuals. The author demonstrated the way in which the structure of the environment in general, and the public transport network in particular, made certain activities possible and other activities impossible. In this way the spatio-temporal reach of an individual can be defined, which Lenntorp originally modeled graphically in the form of a prism (see Figures 4.2 and 4.3). A simulation model was then developed to show what the possible trajectories of an individual are, according to his environment and his capacity of movement. One of the objectives is to "graphically represent and quantitatively determine the part of a person's environment that is physically accessible, i.e. its physical reach. Another objective is to indicate how the quantitative determining factor of the physical reach can be applied and what specific results we can get" [LEN 76a].

4.4.1.1. *General features of the simulation model*

The model, called *Program Evaluating the Set of Alternative Sample Paths* (PESAP), is based on the principle that a population can be studied in terms of spatio-temporal trajectories. The simulation produces information on the possibility of the trajectory of a person by taking into account the constraints of his activity program and of those imposed by the environment: the locations of the stations considered, their opening hours and the spatio-temporal structure of the public transport network.

The user of the simulation model chooses a type of activity program by defining the constraints of departure and arrival times at the point of retreat (dwelling), the places that have to be visited and the possible hours for these visits according to the planned activities (work, public service visit, etc.). The daily prism of the individual is thus defined and we examine how many program combinations exist from each place of the considered environment. The environment is modeled using a georeferenced grid where each cell corresponds to a surface determined according to the needs of the study. Finally, the movement of the individual is considered according to the mode of transport used, its speed and the use of the road network.

4.4.1.2. *The application of Karlstad*

This study, which was done within a government program on the question of public transportation, is described by Lenntorp in a collective publication which attempts to explain the role of time and space in the structuring of society and the environment [LEN 76b]. The objective of the study is to develop an evaluation method of the current and future capacities of the public transport system from the user's point of view in the city of Karlstad[3]. In order to use the PESAP model, the following activity program has been chosen: departure from home – leaving child at daycare – traveling to the Post office – work – pick up child from daycare – return home. This choice is justified insofar as this type of program corresponds to a type of household in which both parents work and at least one of them depends on public transport. Scheduling constraints on maximum transport and journey time between home and work via daycare are associated with this sequence of activities. The data on hours of operation of workplaces and postal services are taken into account. The environment in which the simulation experience is based is defined and modeled, as shown in Figure 4.6. The grid contains 500x500 m squares where the central point (see Figure 4.7) is the departure and arrival points of some of the simulated trajectories.

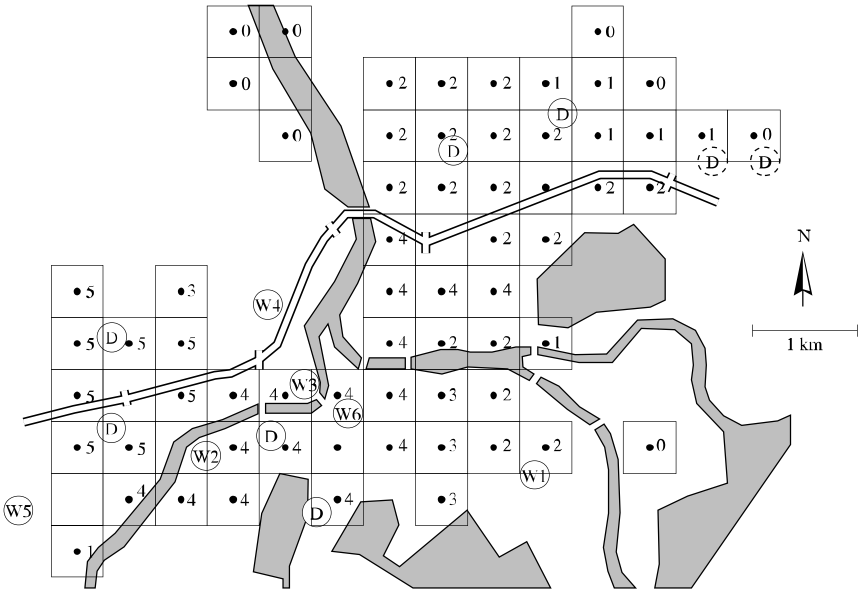
This Figure shows the 62 dwelling locations corresponding that are also the departure and arrival points of the activity programs. The number of daycares corresponds to the actual number of daycares of the town (the two dotted circles are daycares in construction and are considered in the last simulation tests). The six workplaces are the most important workplaces in the city. The grey areas correspond to the highway and lakes. The other lines represent the rest of the transport network.

Figure 4.6. *Location of daycares and workplaces, and representation of the sample of dwellings and public transport network in Karlstad*

---

4 Located at 210 km east of Oslo and 310 km west of Stockholm, the city lays on Lake Väner's north shore and had approximately 72,000 residents and a surface area of 1,000 km^2 in 1975.

The first simulation is based on the existing characteristics of the public transport network. The first results are presented in Figure 4.7. No activity program was able to reach W1, which is too far from the transport networks and daycares, and very few could reach W5. W3 and W6 are the only ones that can be reached from 50 of the 62 points. These areas have the best position in relation to the location of daycares and the transport network. On a global scale, we also see that homes in the West have a more favorable position and that those located in the North and North-East have an unfavorable position in terms of the number of accessible workplaces.



This Figure shows the number of accessible workplaces from the different points, by taking into account a stop at the daycare before and after work. The transport conditions (network and traffic) correspond to the actual situation in the town of Karlstad at the time of study.

Figure 4.7. *Home to workplace travel (W) and stop at daycare (D)*

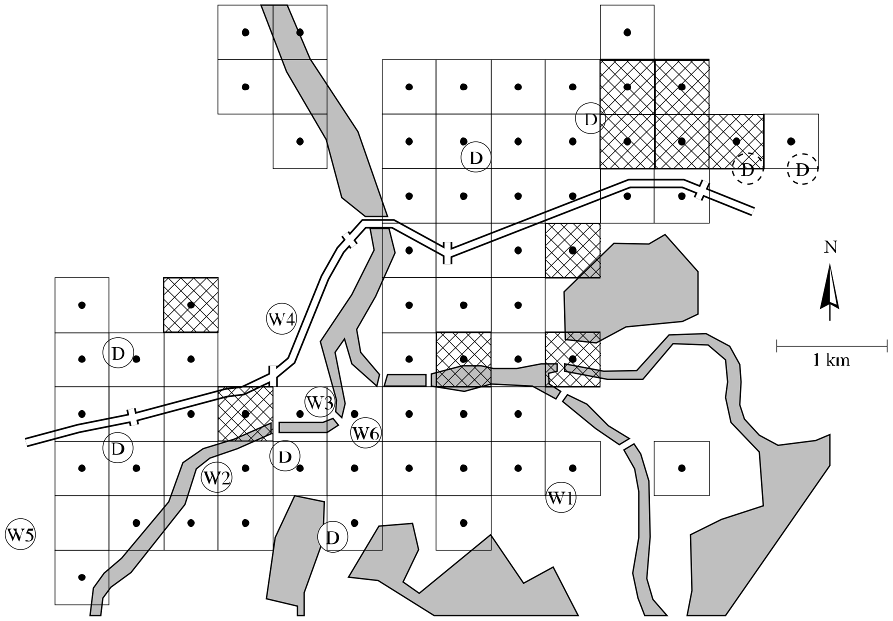
Different scenarios can be tested by changing the parameters of bus traffic on the public transport network. One of the experimentations corresponds to a scenario where approximately 1.5 times more buses are now available to users.

The results of this experiment are shown in Figure 4.8 and lead to the following conclusions: only four workplaces benefit from better accessibility (2, 3, 4 and 6), 10 additional activity programs are possible (for example, departure from North-West cell – daycare – workplace W1, etc.) and the most unfavorable North-East areas experience a small improvement.

These limited results lead us to think that the stop at daycare is undoubtedly the element with the highest constraint in the activity programs and that a much larger increase of traffic is necessary to compensate for this constraint.

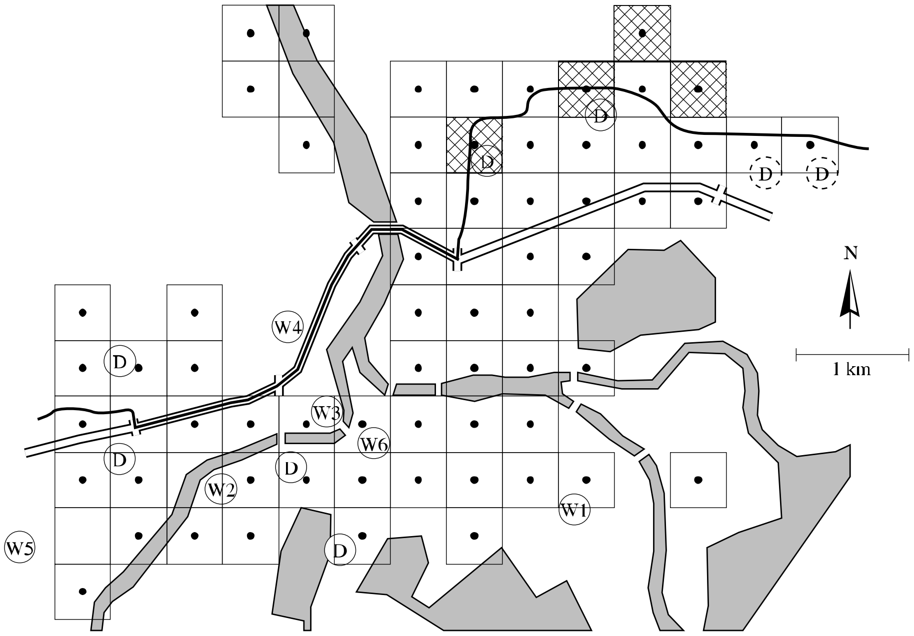
The second experimentation introduces a change in the transport network by adding a circulation line in the north-eastern part of the city (see Figure 4.9). Two additional workplaces are now made accessible and only four new programs can be achieved. In addition, in this new bus network configuration and from different home locations, it is now possible to choose among a higher number of daycares to complete the activity program considered in the simulation. Spatially, the parts which are closest to the new circulation line are those that benefit the most from this change.

Finally, one last simulation has been developed by taking into consideration the daycares in construction, which are surrounded by dotted lines in Figure 4.6. An improvement is noticed for the dwellings located in the area of the new daycares, provided that people in these areas must choose workplace W1 located downtown.



The hatched cells correspond to residential areas from which at least one more workplace is made accessible by the addition of a circulation line on the public transport network.

Figure 4.8. *Change in accessibility due to a higher density of traffic for home to work (W) via daycare (D) travel*



The hatched cells correspond to residential areas from which at least one more workplace is made accessible by the addition of a circulation line in the public transport network that experience an increase of at least one accessible workplace with the addition of a circulation line to the public transport network (as indicated in this figure).

Figure 4.9. *Change of accessibility due to an extended bus network on the home to work (W) via daycare (D) travel*

This application presents in a practical manner the implementation of the initial works of the Lund school in geographical case studies. The originality resides more in the approach proposed to cover a class of traditional problems in geography, i.e. transportation and accessibility, than in the results themselves. The fundamental idea, inherent to the principles of time-geography, is not to lose sight of all the elements involved in a given phenomenon. That is why the question of transportation is not addressed in terms of network and network analysis but in terms of spatio-temporal structure involved in the planning of the individual trajectories of a population. The "transport" question is taken in the context of the general organization of a society in which individual projects must be coordinated.

4.4.1.3. *New implementations and operational methods in time-geographic research*

Lenntorp's study, described above, shows particularly well the characteristics of the first developments in time-geography, which aimed mostly at the creation of an original notation system. The notions developed and used made it possible to refine the "staging" of individual trajectories. The study from Karlstad has made it possible, for example, to clarify the concept of accessibility in the following manner: in given spatio-temporal conditions, some activity programs are possible and others are not. Besides, this accessibility is graphically represented in the form of prisms. Lenntorp [LEN 99] explains this original orientation by the involvement of the Lund research team (*Research Group for Process and System Analysis in Human Geography*) in a government program attempting to develop public service offerings (in full growth at the time) through the planned development of the physical structures of the environment. An approach emphasizing the physical and visual realities of the actions received a favorable response in this context.

Even though these remarkable contributions date back to the late 1970s, they are still a fruitful source of ideas concerning the development of mobility and accessibility measures and modeling. Thanks to the improvement of geo-computational tools (GIS, geovisualization, distributed artificial intelligence such as multi-agent simulation, etc.), there has been a renewal of time-geography research since the mid-1990s with a large spectrum of applications. This new trend has been particularly developed within the field of transportation research with the increasing interest in so-called "activity-based approach". Timmermans *et al.* [TIM O2] show, within a detailed state of the art, how it has led to several modeling approaches.

Among the several studies developed in recent years, we will focus on two particular kinds of implementations of the time-geographic construct:

– *The development of 3-D geovisualization of human activity patterns in space-time.*

Kwan [KWA 99, KWA 04] is one of the leading representatives of this new trend within time-geographical research. She was the first to implement the "notation system" in a 3-D GIS environment. She used, for example, a very rich dataset based on individual-level activity-travel diaries (collected in Columbus, Ohio, USA and in Portland, Oregon, USA) and could visualize each space-time trajectory in its geographical context. Kwan explains that this kind of visualization is of great help in analyzing different groups of population: the main advantage of such a geovisualization analysis is to provide a comprehensive view of complex space-time individual data with a dynamic environment tool that helps the researcher to reveal distinctive space-time patterns and to formulate new hypothesis about accessibility capabilities and space-time use of metropolitan areas. Even if the GIS environment provides a huge capacity of visualization, it is still very difficult to handle the complexity of the scene of a large set of trajectories and we need some complementary methods to reveal the underlying patterns. Starting with the vizualisation of all trajectories, Kwan developed methods in order to standardize trajectories and compare data in a standardized space-time context.

– *The development of individual-based mobility models [BAN 05].*

Recent experiments try to implement the time-geographical constructs by using the paradigm of distributed artificial intelligence. For example, Bellomo and Occelli [BEL 00] developed a model called SimAC (Simulation Accessibility), implemented on the SWARM simulation platform, in order to simulate accessibility based on two main notions: a concept of "activity space" and a notion of urban performance management. The first concept is directly inspired by time-geography: "activity-space" is defined as the functional, spatial and temporal space available to an individual as a result of his participation in activities, in relation to personal and environmental constraints and opportunities.

A more recent research project [BAN 05], MIRO (Model for Intra-Urban Daily Rhythms) seeks to model the global rhythms of a metropolitan area using an individual perspective as a starting-point. This project is connected to the new temporal policies which try to better take into account the individuals' needs in terms of activities, time-budget and travel capabilities. The aim of MIRO is to define a scientific protocol for investigations and simulations that would be make it possible to depict the varying territorial configurations produced by myriad individual trip trajectories. This "bottom-up approach" consists of describing at the individual and local level the constraints faced and some basic behavioral rules, in order to let more complex individual behaviors like activity chronicles and trip chaining emerge, and also more collective behaviors composing urban rhythms. The model is applied to the city of Dijon, France. It is structured in three steps:

– first, a GIS database integrates information concerning the transport supply (networks, capacity and speed, time-schedules for public transport), the residential locations, the description of activity opportunities (opening and closing time, packing capacity of buildings, number of workers, etc.);

– next, a protocol is defined in order to create a synthetic population of agents (this part is inspired by previous works like the TRANSIMS model [BAL 04]). This step aims to generate an "artificial" population of individuals, representative of the population in Dijon, merely described by time-budget activities. The main source of information in such a perspective is data from a travel survey, collected in Dijon, that helps to characterize different types of behaviors in terms of activities, travel mode, social and demographic features;

– finally, a multi-agent-system model is developed, under the MadKit platform [MAR 06]. Agents belong to the so-called cognitive agents. They have a limited knowledge over urban amenities, and they interact locally with their dynamic urban environment and with other agents. Finally, agents have to schedule their daily program of activities and to perform them in a moving urban environment (dealing with traffic conditions, other agents, time schedule of urban opportunities, public transport availability, etc.).

The MIRO model manages to simulate a population of 10,000 agents in Dijon, each agent performing several activities (according to pre-determined activity programs) during 24 hours. When a given agent fails to perform his program during the first simulation period, he is able to change his behavior according to what he learned during the first period. The whole set of activity-programs and travel behaviors are stored, which makes it possible to compare – at the global level – how agents adapt their behavior to the local condition they encounter. From that point, MIRO now seeks to produce a more integrated approach, simulating the dynamic between land use and activity-travel patterns. While it is still in progress, this project shows the inherent complexity we have to face when trying to simulate the "swarming" city.

4.4.1.4. *Partial conclusion*

Parallel to the previous developments, a strong criticism emerged in the 1980s and early 1990s, condemning the time-geographic approach on its "physicalistic" character and finding its analysis of the actions too mechanical, at the cost of the thinking individualities of people, who undergo experiences, feelings and have expectations for the future. In parallel, time-geography has been echoed among sociologists. In particular, Giddens [GID 84] revealed how time-geography contributed to the field of social structures study. He mainly retained the opportunities that time-geography opens for the development of a real *contextual* theory that integrates the spatio-temporal locations of actions within the constitution of social structures. This new scientific context then enabled the appearance of new studies that enriched the bases of time-geography, by drawing on the theoretical references of other social sciences [ELL 00, LEN 04]. The next section presents an example of such a study.

4.4.2. ***Daily lives of women: adaptation strategies in time and space – the Tora Friberg method***

Friberg's [FRI 90, FRI 93] thesis applies to the organization of women's daily lives. The study is done on two different levels: on a macrolevel it establishes the reality of gender segregation in the labor market and on a microlevel it shows through surveys of some women how they organize their daily lives, combining work, housework, family life and spare time.

Friberg's methodology links both levels of analysis by associating in her study the longitudinal approach of time-geography with a theoretical frame based on the analysis of life forms [AQU 92].

4.4.2.1. *From Højrup's life forms to Friberg's three women life forms*

Friberg bases her study on a theory of the life forms developed by a Danish ethnologist, Thomas Højrup [HOJ 89]. Marxism is used as a starting point in the development of this theory; different social classes have different commitments and relations with work, which results in different situations in relation to the other parts of the daily life.

The first life form groups families having their own production tool (farmers, fishermen, tradesmen, etc.), and controlling their own work. The whole family is often involved and the line separating work from spare time is non-existent.

The second form corresponds to a life form of a salaried employee where work is merely a means making recreation possible, for example. Recreation time then gives a possibility for personal fulfillment that work does not provide. Family life is part of recreation even if it is a condition for the salaried work through its reproduction function.

The last life form encompasses the jobs in which a career is possible. The people demonstrate a high interest in their work to a point where it even becomes their main goal. Recreation is used as a way to improve their skills and knowledge. Family becomes in this case a service function facilitating career development.

The main criticism for this theory is that it does not take into consideration gender questions and that it is mainly formulated from the working conditions of men. That is why Friberg revises the typology by directing the context of her study toward women. Three forms of life are studied: the form of life of a salaried employee, a career form of life and an intermediate form.

The life form of a salaried employee keeps a critical relation with the job. The women concerned put high priority on their family, home and housework, combining the features of the ideology of stay-at-home mums with a firm willingness to hold on to a salary.

The career form of life covers women that will give a lot of attention to their work and for whom their family is very important, while housework is left behind.

The third life form comes from the results of Friberg's empirical study where cases that do not fit in either of the two first types have emerged. In this ultimate category are the women that give their work a lot of importance but that do not necessarily want a career. They often work part-time when children are small and the relation with the home and housework is in the middle of the two previous forms.

4.4.2.2. *Relation with time-geography*

The common characteristic of all three women life forms defined by Friberg is that they all strive to combine work, housework and family life. This combination of activities is often difficult to manage because women are confronted with spatio-temporal constraints leading to competition between the different projects. Friberg relies on time-geography to explain these complex and concrete situations of everyday life, using the notions of constraint and project in particular. The analysis of critical situations (temporal conflicts in the activities) highlights recurrences in the contexts favorable to their appearance. The author's interpretation brings out three main features:

– "being sure that is good for the children;

– trying to find sufficient time; and

– coping with the various tasks that must be done" [FRI 93].

These elements favorable to the appearance of conflicting situations in women's activity programs are of a very different nature than those defining capacity, conjunction and power constraints. Maybe in response to the criticism of a too physicalistic characteristic of the original time-geography notions, Friberg emphasizes another type of constraint that she calls *organizational obstacles*. The organizational obstacle illustrates the idea that the scarcity of spatio-temporal resource unavoidably imposes individual *choices* to get out of the *packing dilemma* (too much to do at the same place and at the same time): "An organizational obstacle is the conflict-laden situation that arises when different projects compete with one another about given space in the space-time environment, which is resolved when the individual makes a choice which entails that one project prevails and one or more projects yield ground" [FRI 93]. This type of constraint highlights the real decision capacity of the person giving it a more social dimension.

The contribution of time-geography in this study fundamentally resides in its longitudinal approach, which "follows the individuals" by sequencing and analyzing the *chain* of their activities. Merely measuring the types of projects prevalent in one or another woman life form would not have highlighted the conflicting situations that appear mostly with women, because of the difficult combination of several projects with intermingled and interrelated steps in complex contexts. From this linear approach to women's activity programs, different strategies arise to resolve organizational constraints. Time-geography highlights these variations but does not explain why or how these choices were made. For this, Friberg must rely on the analysis of the forms of life developed in ethnology.

4.5. Conclusion

Time-geography can be considered as the beginning of a powerful semantic modeling from a conceptual standpoint because it guides the reflection by making generalized assumptions on all types of questions on spatio-temporal behaviors. One of the first aims is to explain how people's choices and actions are regulated by space and time constraints in order to evaluate, or even anticipate, the possibilities of completing a project; So, the main goal is not to describe the use of space and time but to show the logics of their usage and their regulators. Thus, time-geography helps us underline constraints that last because they are structural and to build a more solid base for any reflection on space-time.

The graphical aspect of modeling is not as interesting since it is part of a heuristic approach by the "visualization" of the fundamental principles of the semantic model. These graphical tools are the result of a willingness to "show" complex structures whose visual analysis enables us to discover and formulate assumptions on the issues of a world on the move. Their strength resides also in the relatively high level of abstraction (graphical), making them useful for many applications from different domains and disciplines.

However, a criticism can oppose the methodology. It is mainly based on its descriptive character as opposed to explanatory. Often presented as the hallmark of the approach, the 3-D graphics are in reality operationally weak and badly reflect, from this point of view, the conceptual richness of the underlying frame of thought. Since they do not represent a large number of trajectories, graphical models make the process too "exceptionalistic" to be used as a real information processing tool. The precursors of time-geography are refocusing their position today [LEN 99] by stating that the notation system of time-geography is the product of ideas or of an image of the world and not the opposite.

Space is presented in models in terms of infrastructures that are physically stable in time and defined as functional places where people stay to accomplish a given activity. However, nothing explains the preliminary logic of spatial organization of these places or clarifies the meaning of these spaces for people, a meaning that is taken into account in their choices of attendance. The environment and the context of action are presented as "data" on which individuals do not seem to have control. According to Giddens or Buttimer, it leaves little room for the acting and thinking individual, who is capable of major transformation by his actions and social interactions.

This conceptual frame proposes reading the individual trajectories as the result of compatibilities between projects and constraints, needs and resources. However, there is no proposed explanatory outline to clarify the predominant character of certain projects, or to explain the control of some social networks over others in the arbitrations of distribution between needs and resources.

We cannot deny the paradox of an approach that advocates studies oriented on the individual, while confining it to a very functionalist, or even "physicalistic" vision of the world. And we cannot settle for a modeling that aims to grasp the spatio-temporal practices but does not enable real analysis and processing of numerous data.

The lesson that we can learn from these limits is that time-geography proposes a powerful conceptual study grid integrating spatial and temporal dimensions in a given problem, but it cannot overlook the importance of more operational models. It is in this spirit that microsimulation models [HOL 04] (see Chapter 6) or agent-based models are developed in several studies since the 1990s (see also Chapter 7). Another attempt is the pairing of statistical models able to generate typologies of activity chronicles with temporal database systems linked to dynamic cartographic visualization tools [CHA 99]. The complementary use of various models and methods (whether they are linked to spatial analysis or social theories) is undoubtedly the most promising way to pursue in the development of time-geography today. This promise can make us think, to paraphrase a formula from Lenntorp [LEN 99], that we are at the end of the introduction of time-geography.

4.6. Bibliography

[ALO 64] ALONSO W., *Location and Land Use – Towards a General Theory of Land Rent*, Cambridge University Press, 1964.

[AQU 92] ÅQUIST A.C., *Tidsgeografi i samspel med samhällsteori*, Meddelanden från Lunds universitets geografiska institutioner, Avhandlingar 115, Lund, 1992.

[BAI 90] BAILLY A., BEGUIN H., *Introduction à la géographie humaine*, Masson, Paris, 1990.

[BAL 04] BALMER, M., NAGEL K., RANEY B. Large scale multi-agent simulations for transportation applications, *EIRASS Conference* 2004.

[BAN 05] BANOS A., CHARDONNEL S., LANG C., MARILLEAU N., THEVENIN T. "Simulating the swarm city: a MAS approach", *CUPUM'O5 Conference on Computers in Urban Planning and Urban Management*, London, 2005.

[BAU 92] BAUER D., MARESCA B., *Lignes de vie. Méthodologie de recueil et de traitement des données biographiques. Le cas des carrières et trajectoires professionnelles*, Credoc research notebook, Paris, 1992.

[BEL 00] BELLOMO M., OCCELLI S., "Simulating Accessibility by SWARM", in *Applications of Simulation to Social Science,* G. Ballot, G. Weisbuch (eds.), Hermes, Paris, p.141-154, 2000.

[BOU 02] BOULIN J-Y., MÜCKENBERGER U., *La ville à mille temps*, éditions de l'aube, La Tour d'Aigues, 2002.

[CAR 78a] CARLSTEIN T., PARKES D., THRIFT N., *Human Activity and Time Geography*, Arnold, London, 1978.

[CAR 78b] CARLSTEIN T., PARKES D., THRIFT N. (eds.), *Timing Space and Spacing Time*, E. Arnold, London, 1978.

[CAR 82] CARLSTEIN T., *Time, Resources, Society and Ecology*, Allen and Unwin, London, 1982.

[CHA 93] CHARDONNEL S., La géographie des "espaces-temps", étude spatio-temporelle de la vie quotidienne d'une famille suédoise, T.E.R., Grenoble, Alpine Geography Institute, 1993.

[CHA 99] CHARDONNEL S., Emplois du temps et de l'espace. Pratiques des populations dans une station touristique de montagne, Geography Thesis, University of Grenoble, 1999.

[CLA 84] CLAVAL P., *Géographie humaine et économique contemporaine*, PUF, Paris, 1984.

[CLA 99] CLAVAL P., "Des aires culturelles aux réseaux culturels", *L'approche culturelle en géographie, conference for APHG, Caen regional*, Caen, 1999.

[DIM 99] DI MEO G., "Géographies tranquilles du quotidien. Une analyse de la contribution des sciences sociales et de la géographie à l'étude des pratiques spatiales", *Les Cahiers de Géographie du Québec*, vol. 43, no. 118, p. 75-93, 1999.

[DUM 00] DUMONTIER F., PAN KE SHON J.L., "Enquête Emploi du temps 1998-99. Description des activités quotidiennes", *Insee Résultats* no. 693-694, *Consommation – Modes de vie* no. 101-102, Insee, Paris, 2000.

[ELL 89] ELLEGÅRD K., *Akrobatik i tidens väv*, Choros Kulturgeorgrafiska insitutionen, Handelshögskolan vid Göteborgs universitet 2, Göteborg, 1989.

[ELL 91] ELLEGÅRD K., ENGSTRÖM T., NILSSON L., *Reforming Industrial Work - Principles and Realities: In the Planning of Volvo's Car Assembly Plant in Uddevalla*, Arbetsmiljöfonden, Stockholm, 1991.

[ELL 96] ELLEGÅRD K., "Restructuring automobile industry in Sweden – the emergence of a reflective production system", *European Planning Studies*, vol. 4, no. 4, 1996.

[ELL 99], ELLEGÅRD K., PATER B.D., "Introduction: the complex tapestry of everyday life", *Geojournal* 48: 149-153, 1999.

[ELL 00] ELLEGARD K., PATER B., "Introduction: the complex tapestry of everyday life", *Geojournal* 48, 2000.

[FOR 69] FORRESTER J.W., *Urban Dynamics*, MIT Press, Cambridge, 1969.

[FRA 98] FRÄNDBERG C., *Distance Matters: an inquiry into the relation between transport and environmental sustainability in tourism*, Humanekologiska skrifter nr.15. Avdelningen för humanekologi, tvärvetenskapliga studier av människans villkor, Götebergs Universitet, Göteborg, 1998.

[FRI 90] FRIBERG T., *Kvinnors vardag. Om kvinnors arbete och liv. Anpassningsstrategier i tid och rum*, Meddelanden från Lunds universitets geografiska insitution, avhandlingar 109, Lund University Press, Lund, 1990.

[FRI 93] FRIBERG T., *Everyday Life. Women's Adaptive Strategies in Time and Space*, Byggforksningsradet, Stockholm, 1993.

[GID 84] GIDDENS A., *The Constitution of Society*, Polity Press, Cambridge, 1984.

[GOU 92] GOULD P., "Epidémiologie et maladie", in *Encyclopédie de géographie*, Bailly A., Ferras R., Pumain D. (eds.), Economica, Paris, p. 969-990,1992.

[GRA 79] GRAFMEYER Y., JOSEPH I., *L'école de Chicago – Naissance de l'écologie urbaine*, Champ Urbain, Paris, 1979.

[HAG 53] HÄGERSTRAND T., *Innovationsförloppet ur korologisk synpunkt*, Meddelanden från Lunds universitets geografiska insitution, avhandlingar XXV, Lund, 1953.

[HAG 70a] HÄGERSTRAND T., "Tidsanvändning och omgivningsstruktur", in *S.O.U. 1970:14 Urbanisering i Sverige, en geografisk samhällsanalys*, Bilagedel 1, Balenserad regional utveckling, Stockholm, 1970.

[HAG 70b] HÄGERSTAND T., "What about people in regional science?", *Papers of the Regional Science Association*, 24, p. 7-21, 1970.

[HAG 81] HÄGERSTRAND T., "Interdépendances dans l'utilisation du temps", *Temps Libre*, 3, p. 53-68, 1981.

[HAG 82] HÄGERSTRAND T., "Diorama, Path and Project", in *Tidschrift voor economische en sociale geografie*, vol. 73, no. 6, 1982.

[HAG 85] HÄGERSTRAND T., "Time-Geography: focus on the corporeality of man, society, and environment", in *The Science of Praxis of Complexity*, The United Nations University, p. 193-216, 1985.

[HAG 91] HÄGERTRAND T., *Om tidens vidd och tingens ordning Texter av Torsten Hägerstrand*, articles assembled by Carlestam G., Sollbe B., Byggforskningsrådet, Stockholm, 1991.

[HAR 39] HARTSHORNE R., *The Nature of Geography*, Association of American Geographer, Lancaster (Pa.), 1939.

[HOJ 89] HØJRUP T., CHRISTENSEN L.R., *Sturukturel livsformsanalyse*, Nordisk tidskrift för folkelivsforkning 37, Nord-Nytt, 1989.

[HOL 04] HOLM E., HOLME K., MÄKILÄ K., MATTSON-KAUPPI M., MÖRTVIK G, "The microsimulation model SVERIGE: content, validation and applications", SMS, Kiruna, Sweden (www.sms.kiruna.se).

[KJE 95] KJELLMAN C., *Prokrami-modellen: Ett trovärdigt behandlingsalternativ*, Kulturgeografiska Institutionen, Lund, 1995.

[KJE 03], KJELLMAN C., "Ta eller få plats ? Studier av marginaliserade människors förändrade vardagsliv, Meddelande från Lunds Universitets Geografiska Institutioner", avhandlingar nr 148, Sociologiska institutionen, Lund, 2003.

[KWA 04] KWAN M.P., "GIS methods in time-geographic research", *Geografiska Annaler*, 86B(4), p.267-280, 2004.

[LAN 97] LANGLOIS A., PHIPPS M., Automates cellulaires. Application à la simulation *urbaine*, Hermès, Paris, 1997.

[LEL 99] LELIEVRE E. (think tank), *Biographies d'enquêtes, Bilan de 14 collectes*, Ined-Puf, Méthodes et Savoirs, Paris, 1999.

[LEN 76a] LENNTORP B., *Path in Space-Time Environments. A Time-Geographic Study of Movement Possibilities of Individuals*, Meddelanden från Lunds universitets geografiska institution, Avhandlingar LXXVII, Lund, 1976.

[LEN 76b] LENNTORP B., A time-space structured study of the travel possibilities of the public transport passenger, Report och Notiser 24, Lunds Universitets Kulturgeografiska Institution, Lund, 1976.

[LEN 78] LENNTORP B., *A Time-Geographic Simulation Model of Individual Activity Programmes*, in Carlstein T., Parkes D., Thrift N. (eds.), *Timing space and spacing time*, E. Arnold, London, 1978.

[LEN 98] LENNTORP B., "Orienteringsanvisning i ett forskningslandskap", in *Svensk Kulturgeografi: En exkursion inför 2000-talet*, Gren M., Hallin P.O., Studentlitteratur, Lund, p. 67-85, 1998.

[LEN 99] LENNTORP B., "Time-geography – at the end of its beginning", *Geojournal*, Kluwer Academic Publishers, 48, p. 155-158, 1999.

[LEN 03] LENNTORP B., "The drama of real-life in a time-geographic disguise", 6èmes Rencontres de TheoQuant, 2003.

[LEN 04] LENNTORP B., "Path, prism, project, pocket and population: an introduction", *Geografiska Annaler*, 86B(4), p.223-226, 2004.

[MAR 79] MÅRTENSSON S., *On the Formation of Biographies in Space-Time Environments*, Meddelanden från Lunds universitets geografiska institution, Avhandlingar LXXXIV, Lund, 1979.

[MAR 06] MARILLEAU N., "Méthodologie, formalismes et outils de modélisation-simulation des systèmes complexes de mobiles: application à la mobilité géographique", PhD Thesis, LIFC, University of Franche-Comté, 2006.

[OBE 89] ÖBERG S., HOLM E., MÄKILÄ K., *Tidsgeografisk handlingsteori: att bilda betingade biografer*, Geografiska institutionen Umeå universitet, Umeå, 1989.

[PAR 80] PARKES D., THIFT N., *Times, Spaces and Places*, John Wiley, New York, 1980.

[RYD 92] RYDENSTAM K., “Så använder vi tiden”, in *Välfärds Bulletinen*, no. 4, Statistiska Centralbyrån, Stockholm, 1992.

[THR 77] THRIFT N., “An introduction to time-geography”, *CATMOG*, no. 13, London, 1977.

[TIM 02] TIMMERMANS H. ARENTZE T, JOE C-H, “Analysing space-time behaviour: new approaches to old problems”, *Progress in Human Geography,* 26:2, p. 175-190, 2002.

[VAN 99] Van der KNAAP W., “Research Report: GIS Oriented Analysis of Tourist Time-Space Patterns to Support Sustainable Tourism Development”, Tourism Geographies, vol. 1, no. 1, 1999.

Chapter 5

The Process of Spatial Diffusion and Modeling Change

5.1. Introduction

Diffusion processes occupy a place of choice in the expression of dynamics affecting spatial systems. Certain characteristics of these processes justify the importance given to them in the spatial change component inventory as well as in a presentation of the methodologies explaining their features. We will also add that the studies on spatial diffusion of innovations can be considered as precursor studies paving the way for the contemporary attempts to model spatial change.

A brief terminology reminder will be useful in introducing the chapter. Generally, by diffusion we mean the action, and the result of the action, of propagating an object in a homogenous way or transmitting the object or a phenomenon in a system whatever the propelling power. At the end of the process, the system is saturated. Its absorption capabilities are exhausted. The system invested by this action tends to go from one state to another. The notion of diffusion is therefore mobilized every time we study processes that involve movements of matter, products, people, intangible goods, practices or ideas for penetrating a system. Due to this, the diffusion processes are addressed by physical as well as biological and human sciences.

Social sciences in general and geography in particular often associate diffusion and innovation (see Box 5.1). In its most trivial sense, innovation simply represents

---

Chapter written by Thérèse SAINT-JULIEN.

the newest thing (material or intangible) that, when it appears for the first time, is adopted by a person or a group. We have also studied the diffusion of contagious illnesses as well as of objects (such as, for example, a form of housing, a mode of transport, a new industrial product, an agricultural technique) and of intangible goods (for example, a type of economic development, a religious or a linguistic practice, a social model, etc.) which are all new things appearing at different places where they were not present before. In its strongest sense, innovating means introducing something new into an established thing, which can transform it. Thus defined, the innovation generates, by propagating, irreversibilities in the evolution of the environment that it penetrates [SCH 34]. The more complex it is, the more its diffusion will become a determining component in the transformation of the environment in which it appeared due to the growth of the induced effects that its adoption generates. Even if diffusion processes can be located and simulated for any new phenomenon that spreads over a system, the interest that is consistently given to these processes stems from the structuring effects that some of them hold.

The spatio-temporal dimension of the diffusion process understood through the notion of spatial diffusion is considered when the diffusion is done in spatial systems. The elements of the system concerned by the innovation are identified by, among other things, their location. The focus of this chapter is the notion of spatial diffusion which covers all the processes competing for the movement and the migration of the innovation in the geographical space and for the effects that these movements generate in that space.

5.2. The manifestations of diffusion in space

Certain very general regularities of spatial diffusion processes justify the modeling examinations that have been done to them. The empirical observation of the diffusion processes has enabled us to locate two major forms of spatial diffusion. The first form corresponds to a progressive *spread* of the zone covered by the propagation that, from the transmitting source, seems to occur by degrees and spread like wildfire without diminishing the intensity of the phenomenon's penetration in the areas reached in the first place. It is that way, for example, in the diffusion of new equipment, new technologies and new lifestyles in a geographical system. The second form is a diffusion with *migration* where the places initially reached are abandoned as the propagation is deployed. The receiving potential for innovation is progressively destroyed by the new phenomenon that appears. The epidemiologists know this type of diffusion quite well. We find it in all space occupation processes, when the forms of ground occupation correspond to onsite exploitation of non-renewable resources. The front line of exploitation moves from one site to another as the sites become impoverished, ruined mushroom towns near closed mine shafts are quickly rebuilt elsewhere on new exploitation sites, agricultural plantation front taking over new spaces leaving behind devastated soil. In a large number of cases, both forms can be combined and the expansion then corresponds to a movement of the center of gravity of the diffused elements distribution.

5.2.1. *Elements and levels of approach of a spatial diffusion process*

A spatial diffusion process is associated with an identified *object* destined to propagate in space. We have mentioned that this object can be material. It can also be immaterial: a cultural practice, a political, judicial or administrative model, a regulation, a mode of living, a technology, a knowledge, an architectural model, an economic development model, etc. The innovation is diffused from a *transmitting source*, i.e. a place from which there is a "push" effect, whose strength comes from the union of actors, initiators of the movement. We presume that all the locations do not have the same transmitting source vocation, since this initial propagation strength is a vital determining factor of the diffusion. The role played by large and very large cities, by main economic sources, by all major centers, in and around which efficient contact and circulation networks are structured, is a determining factor.

In order to launch, the spatial diffusion process must be able to depend on *potential adopters*. Individual adopters are associated with certain diffused objects. It works that way for the diffusion of consumer objects, personal services, lifestyles, or demographic behaviors (contraceptive practices, for example, where the potential adopters are people of childbearing age). Conversely, certain innovations concern corporate bodies including companies, institutions (technical innovation, organization mode, management mode, etc.), or even territorial entities such as central or regional governments, cities, local governments, in the case of the diffusion of collective equipment, of a system of representation, of a mode of administration, etc. These particular destinations contribute to the introduction of characteristics in the processes, whether they are relations between transmitters and adopters or their modes of spatial registration and especially geographical levels to which diffusion can be seen.

Box 5.1. *The emergence of the diffusion notion in social sciences*

In the 20th century, the diffusion of innovations, the propagation of changes in one environment, their transition from one environment to another during these movements were permanent concerns within the social sciences body.

Sociology studies have quickly addressed the diffusion problem by a questioning on social change. These studies have particularly emphasized the distinction between endogenous and exogenous change processes. In the explanation of social changes, the first processes have been supported by the individualistic methodology in which, in the Weber tradition for example, the reconstitution of the motivations of individuals affected by the change is essential. In a more global social perspective, almost completely influenced by the works of Durkheim, we have turned to the identification of the external forces to explain the change. This duality of components has been taken up again somewhat explicitly in the studies that have integrated spatial dimension of the propagation.

The works of Schumpeter [SCH 34] on the role of innovation in economic evolution has affected the whole questioning of the 20th century. According to this author, the production of a new product, the introduction of a new method, the opening of a new market or finding a new energy source or raw material are not isolated facts. Each of these facts must be interpreted as the revelation of a larger combination of qualitative changes introduced by company owners. These combinations can be considered as innovations when their propagation in an environment generates irreversibilities in the evolution of this environment. The innovations are not only inventions and even less fads. For Schumpeter, the diffusion of innovations modifies the dynamics of systems in which it operates. Several points are still being debated. Besides the general concern still not addressed of the exogenous or endogenous character of the emergence of the innovation and its impulsion, the discussions focus mostly on the nature and the role of technological innovations in the appearance of economic growth cycles [FRE 85, HAL 88, MEN 79] and on the regularity of these cycles. They also involve temporalities of diffusion processes, whether we are talking of the phases of take off or those of expansion. And finally, the debate is also about the role played by organizational, institutional and social innovations as well as by the government in the diffusion of technological innovations.

The spatio-temporal dimension of the diffusion processes has been specifically introduced in the anthropology, archeology and geography works. In the 1930s the notion of cultural sphere has been introduced in order to explain the diffusion of cultural characteristics. Later, believing that we could locate time by only studying space, the "diffusionism" trend was developed [BOA 40]. Similar cultural elements are identified within different groups. This similarity can certainly refer to the unit of man, but it also suggests borrowings from other cultures, each culture having a certain number of relatively independent characteristics which can migrate from one culture to another. Starting with the complexity principle which says that we always consider a group of cultural characteristics, chance could not explain that the same characteristics appear in different places. If, in addition, these characteristics appear in continuous spheres, the diffusionism theory claims that it is very likely that we have propagation from the center toward the periphery. We can then presume that the characteristics with the largest diffusion area are the oldest. This approach, which deliberately but somewhat schematically integrated the spatio-temporal dimension of the cultural phenomenon, has stirred up much criticism. There were objections that the diffusion is not sequential, especially from the borders rather than from the center, or that in the transition from one culture to another, the borrowings are modified in such a way that they become unrecognizable and that the diffusion results less in borrowings than in a creation. Lévi-Strauss [LEV 58] has proposed that the notion of similarity, at the base of this debate, should be replaced by the notion of affinity, since the diffusion is not only carried out by borrowings or refusals, but also by antithesis.

In the first half of the 20th century, the debate raised by the diffusionism theory has been present throughout the questioning that was developed in geography concerning the notion of lifestyle [GEO 51, SOR 48, VID 11] and was never completely explained. In the 1960s, Gourou [GOU 68] and his students introduced production technique notions and territorial control technique notions in the tropical world with questions on the internal and external forces of change of these techniques and the forms of diffusion of these forces which have reappeared in this field. Archeology has also been confronted with similar questions [VAN 89]. A nomothetic approach of spatial diffusion was actually introduced by Hägerstrand [HAG 52] who has emphasized the existence of temporal and spatial regularities in the innovation diffusion processes, with the help of several case studies. These regularities have led him to pave the way to modelings of these processes and to revive questioning on the role of these processes in the structuring and dynamics of geographical spaces.

Finally, in the same vein as the works of Schumpeter, the theory of growth centers has emphasized the role of driving companies in economic development and particularly in regional development [PER 57]. The expansions brought to this theory by Friedmann [FRI 69] and Boudeville [BOU 72] have helped to clarify the way in which material and cultural innovations, which are bases of economic development, appear in and are diffused by urban networks. By doing this, both authors have focused on the spatio-temporal conditions of this diffusion, thus enriching the geographical approach of diffusion processes.

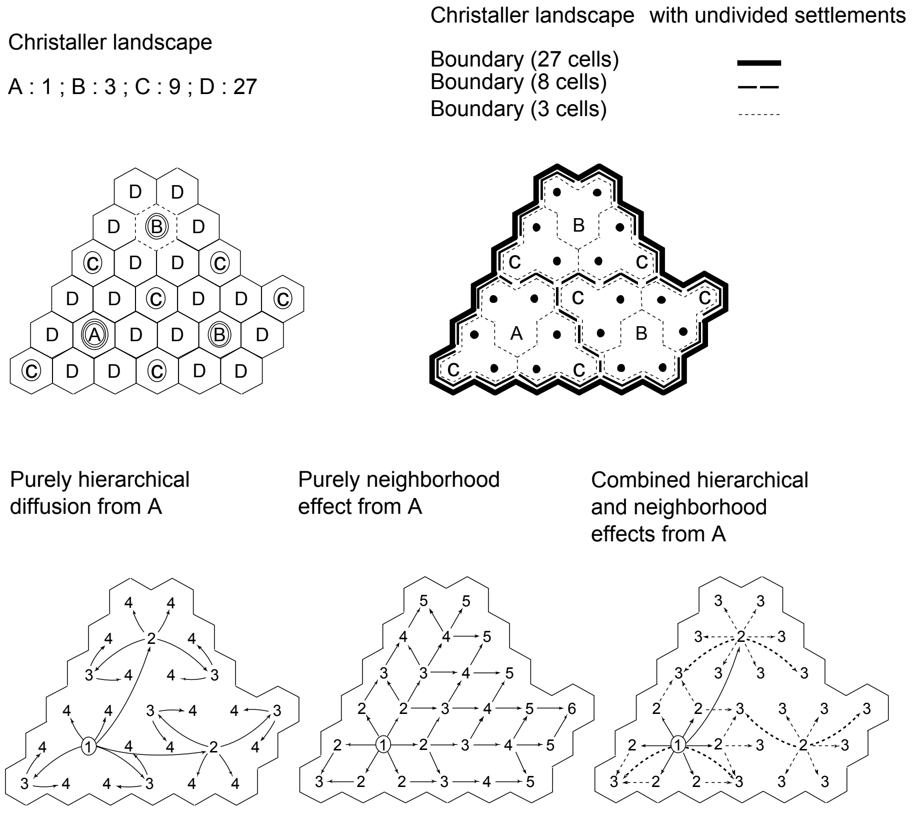
The level of diffusion should not be confused with the level at which the process will ultimately be studied. This process goes through the type of spatial mesh by which the elements involved in the diffusion, or that may become involved in it, will be analyzed. The choice of the level is obviously never independent from the nature of the studied process and therefore from the type of adopters involved. However, the verification modes of the adopted phenomenon throughout time impose a level of approach that is not always the best suited for the process studied. In their study on spatial diffusion of contraceptive methods in Great Britain in the 19th century, Bocquet-Appel and Jakobi [BOC 97] started with an observation available at county level to define the diffusion of these practices. The authors considered that this level imposed by the sources was much too rough to understand the moment when, in each place, contraception is adopted. The map of Great Britain has been subdivided in pixels. With the help of the "krigeage" method, starting with the values of the sampled points from the counties' administrative centers, values have been estimated in the middle of the pixels, thus providing a representation of the distribution of the demographic variable almost continuously. We can also find ourselves in the case where a particular geographical level is not obvious *a priori*, whereas one of the objects of the study is precisely to find the relevant geographical level in order to understand the issues of the process. We can then legitimately explore the same process at different geographical levels. That is what Guerreau [GUE 81] has done, for example. In order to define the characteristics of the diffusion of the mendicant orders in medieval France, he consecutively examined their propagation at city and department level, which brings a new light on the process studied.

5.2.2. *Distances and propagation channels*

The propagation takes place as long as transmitters and adopters have been connected; it is the direct result of the conditions of spatial interaction between transmitters and adopters, of the characteristics of the channels of diffusion. The *direct contact* between transmitters and adopters is a privileged channel. Generally, the probabilities of contact decrease with *distance*. For a given object, propagation in space is based on the distance between transmitters and adopters. The role of distance is complex as the physical distance is not necessarily the only one involved. Distance-time, distance-cost, social, economic or cultural distance and perceived distance can all occur to explain a relative proximity or a relative separation of transmitters and adopters. In addition, at the subsequent stages of a process, these different types of distance either play successive roles (word of mouth, then medium distance communication, etc.) or combine at the same moment. Finally, all things being equal in the nature of the distance, the reduction of the probability of contact with separation will have a different rule to follow. The most common functions to describe this reduction of interactions with distance are either negative exponential functions or power functions or a combination of the two. For a listing of the different distance types and common functions that take into account the reduction of interactions with distance, see Pumain and Saint-Julien [PUM 97, PUM 01].

A large number of empirical studies describe a process that transmits phenomenon through a regular graduation of order, classes or hierarchies (cascade diffusion). These studies have shown how the *hierarchical organization of space* has played an important part. The observation has systematically reminded us that the descending channel of the urban hierarchy was favored in a large number of diffusion processes with, within this hierarchy, highly unequal relative positions of the centers (see Figure 5.1). We have emphasized the vital role of the larger cities as instigation centers. As long as these cities concentrate large masses of potential adopters in a large proximity, they are well positioned to make sure that the adoption of the innovation can be quick and massive and can then quickly acquire transmitting capacities placing them in a favorable hierarchical position to instigate the diffusion toward less important centers. All distances being equal, the diffusion tends to go from large centers to small ones, following the descending direction of the urban hierarchy. In the same principle that all distances are equal, the diffusion will be faster between large centers than between small centers. We notice that the hierarchical propagation channel is often predominant in a small and average scale, whereas it tends to be displaced by the direct contact channel on a large scale. Therefore we can notice that, in a large number of processes, the propagation that follows the urban hierarchy channel and the propagation by degrees combine in a different way according to the geographical levels.

The diffusion in France of high level soccer teams is a good illustration of this combination ([RAV 98], see Figure 5.2). The diffusion started from the neighboring countries of France. High level soccer started propagating by degrees in peripheral regions of the French territory and in some very large cities. This sport was only later diffused on average cities within the territory.



Numbers 1,2,3,4,5,6 indicate the successive stages of diffusion.

Figure 5.1. *Patterns of diffusion according to [HAGG 77]*

In the case of the diffusion of certain goods and services, the territorial logistics of companies must be considered. Spatial diffusion of the good or service can be orchestrated by counting on the conquest of the proximity market; regional diffusion is preferred to establish the product and try out expansion strategies in a known market. Conversely, the choice can be made for a certain category of centers in order to diversify from the outset the experience acquired and, by using many transmitting centers, to better prepare for a future expansion on a large and regionally diversified area. However, the choice may also simply be to develop as well as possible the "natural" channels of the spatial diffusion by adjusting the market conquest strategies according to them, which are presumed able to accentuate their efficiency.

These strategies must also take into account the territorial characteristics of the competition systems.

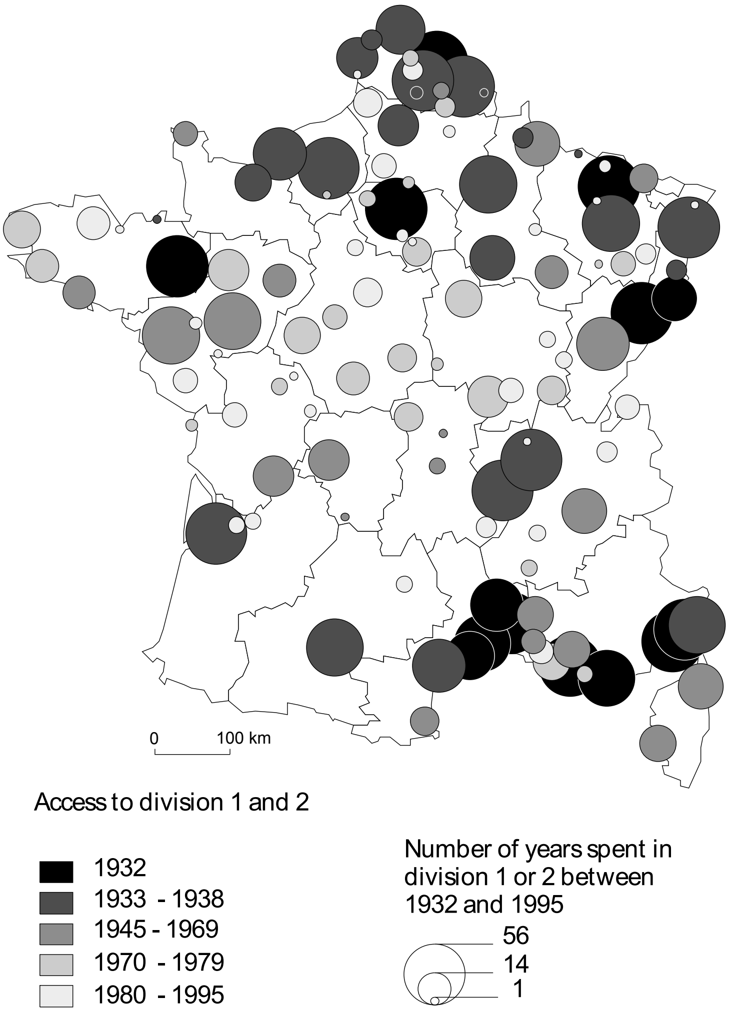
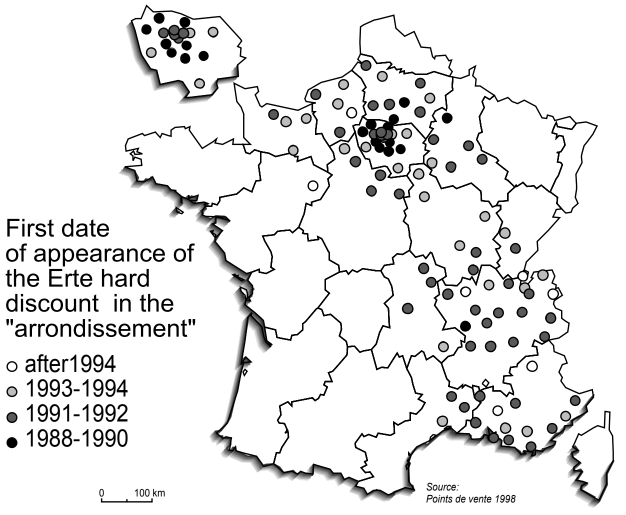
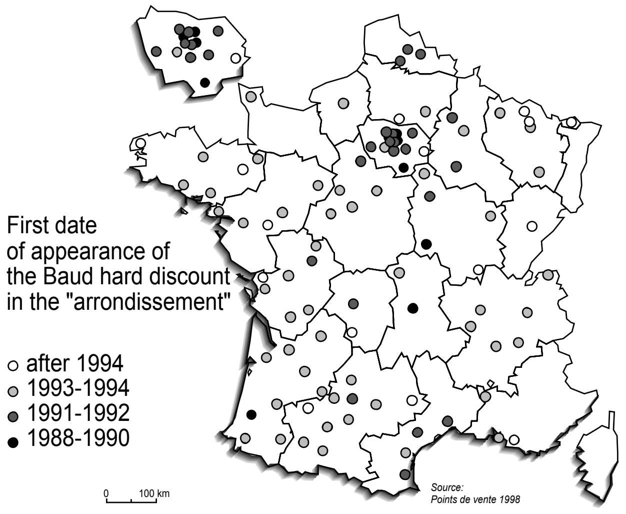


Figure 5.2. *Diffusion process combining hierarchy and neighborhood effects: the diffusion of high level soccer (1932-1995) according to [RAV 98]*

Erte group : diffusion from the Ile-de-France



Baud group : diffusion from the Ile-de-France



according to [SAI 99]

Figure 5.3. *Spatial diffusion in a competition context*

Starting from the Ile-de-France, the spatial diffusion of supermarkets from two groups of large distribution hard discount retailers on the French territory shows the characteristics of these strategies ([SAI 99], see Figure 5.3). Erte group, which is the first one to launch the conquering strategy of the national market, favors regions with the highest urban population density. It avoids northern and northeastern regions, however, those regions having seen the deployment of supermarkets of the hard discount type initially implemented in Germany. So Erte directs its strategy mainly towards regions in the center-east and south-east. Afterwards, the diffusion of the Baud group is done by preferring west and southwest regions, until now not really touched by this type of distribution. The diffusion therefore simultaneously considers the location of the transmitting source, the previous advances from competing groups on the territory and the propagation strategies inherent to each group.

On the whole, the *structures of geographical space* and the technical level of networks linking transmitters and adopters at the time of diffusion are determining factors. Communication networks and all the connection chains are always diffusion accelerators. It goes the same way with social networks [DEG 94] or again, in the propagation of epidemics, with all types of connections such as water, food chains, etc., which are efficient propagation supports. Conversely, barriers stop or thwart spatial diffusion. As with distances, these can be varied in nature (physical, political, social or cultural). Their identification is relative to the process considered, at the observation level and at the moment when it is taken into consideration. Modelings of diffusion processes are used for integrating unequal permeabilities of barriers that can either stop the diffusion or, more often, slow it down or deflect the trajectories.

5.2.3. *Spatial diffusion in time*

We have been able to locate the regularities from which the diffusion processes follow in time. If we enter time as an abscissa and as an ordinate the rate of penetration of the innovation in the spatial system, the rates are ordered in an S curve (see Figure 5.4a). These curves, in the well known form of logistics functions, have two limits corresponding respectively to the number of adopters already hit at the start of the process (X_0) and to the total number of adopters that could receive the innovation during diffusion time ($X_T = N$). The discrepancies between the curves corresponding to the diffusion of an object in different spatial systems result in the differences in penetration rates reached at initial state and especially in the disparities of growth rates, which are indicators of the environment's reaction capacity to the adoption of the innovation.

We can associate with these curves the main steps of spatial diffusion processes (see Figure 5.4b). At the initial stage, the diffusion is relatively slow and the first transmission center(s) are isolated. At the diffusion stage, growth that is still very active at the center is much faster in the periphery, the regions involved in the diffusion multiply and the propagation area quickly expands. At the condensation stage, the process slows down: growth at the center and growth in the periphery have similar rhythms. With saturation, the process is ending. Almost all the potential receiving places have been reached.

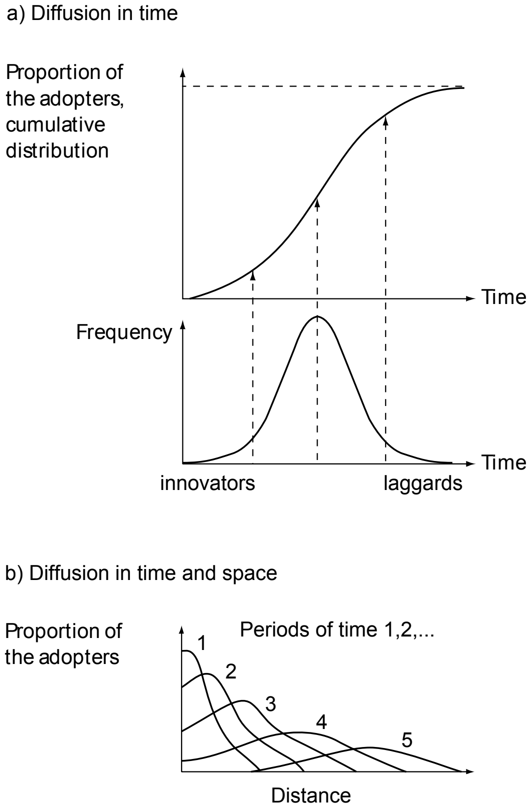


Figure 5.4. *Main spatial diffusion steps*

5.3. Simulating a spatial diffusion process: Hägerstrand's pioneer approach

Hägerstrand's [HAG 67] works on spatial diffusion of innovations have given a decisive impetus to the modeling of these processes. Even though methodological improvements have since been done, the major principles at the base of the approach have been put in place by him. Beyond these technological evolutions that have disrupted simulation methods, we are still struck by certain continuities of the retained approaches. The characteristics of the basic spatial diffusion model proposed by Hägerstrand can be summarized in the following way.

5.3.1. *A probabilistic model*

For a given process, we define all over the space concerned the probabilities of contact between transmitters and adopters (see Table 5.1). These probabilities correspond to the choice of a spatial interaction function that takes into account at the same time the phenomenon whose diffusion is to be simulated, and the general spatial context. All these probabilities, defined in a grid based on a transmitting cell, constitute the field of contact. The central cell of the grid corresponds to the location of the transmitting source at time T_n. The probability of contact y is then higher than elsewhere. The probabilities decrease from this central cell toward the peripheral cells according to the function retained to incorporate the effect of distance in the model.

a) Contact field: probability of contact

0.0096	0.0140	0.0168	0.0140	0.0096
0.0140	0.0301	0.0547	0.0301	0.0140
0.0168	0.0547	0.4432 ♣	0.0547	0.0168
0.0140	0.0301	0.0547	0.0301	0.0140
0.0096	0.0140	0.0168	0.0140	0.0096

Probability of receiving the innovation from a transmitter located in the central cell (♣).

b) Mean information field

0-95	96-235	236-403	404-543	544-639
640-779	780-1,080	1,081-1,627	1,628-1,928	1,929-2,068
2,069-2,236	2,237-2,783	2,784-7,214 ♣	7,215-7,761	7,762-7,929
7,930-8,069	8,070-8,370	8,371-8,917	8,918-9,218	9,219-9,358
9,359-9,454	9,455-9,594	9,595-9,762	9,763-9,902	9,903-9,999

The mean information field is built from the contact field. The probability intervals of each cell are represented in ten thousandths of the total of the field. At each stage of the simulation procedure, it will be centered successively on each of the transmitting cells.

The mean information field is only another expression for contact field. It makes it possible to designate the cell reached during the interval T_n-T_{n+1} *by draw from a table of random numbers.*

Table 5.1. *Contact field in the hypothesis of a homogenous distribution of potential adopters (according to Hägerstrand [HAG 67])*

5.3.2. *The rules of the basic model*

5.3.2.1. *Diffusion in a homogenous space*

The zone studied is a homogenous space on which the potential adopters are uniformly distributed (see Table 5.1). It is divided by a regular grid. Each cell has the same number of potential adopters.

The diffusion process time is itself divided in equal intervals. At an initial chosen moment T_0 the source cells are located, i.e. all those that have already received at least once the new phenomenon and have become able to transmit the innovation to another unit of the cell or to units located in other cells.

From a source cell, the message can only be transmitted once during each discrete time period. The transmission is done through the contact between two cells only. The probability that the cells receive source information depends on the defined contact field. The adoption is executed as soon as a message is received. A cell receives the message at time T_1 and retransmits it at time T_2.

The information received by cells that have already adopted the innovation does not affect the situation. The messages falling beyond the borders are considered lost.

At each time interval, the average field of information is successively centered in each source cell. The location of the contact field cell receiving the innovation is determined by a random procedure.

Diffusion can be interrupted at any stage of the process.

When all the cells have received the innovation, the diffusion process ends. After a large number of draws, we get the most probable form of the spatial expansion of the diffusion process.

5.3.2.2. *Diffusion in a heterogenous space*

The basic model has gone through several enhancements in order to adapt to specific processes that have been studied and modeled. We can particularly mention the integration of heterogenity of forms of the geographical space (see Table 5.2):

– the uneven distribution of potential adopters is taken into account in the contact probability between the transmitting and the receiving cells. We generally resort to a weighting of probabilities so that the probability that innovation is accepted in a cell increases with the number of the cell’s potential adopters, and then decreases as the cell gets closer to saturation;

a) Distribution of potential adopters N_{ij} around cell x_3, y_3.

	y_1	y_2	y_3	y_4	y_5
x_1	6	9	4	12	1
x_2	1	3	2	2	16
x_3	8	23	4	7	2
x_4	5	21	9	12	6
x_5	2	6	11	26	5

The distribution of potential adopters varies according to the transmitting cell used.

b) Contact field

0.01	0.014	0.017	0.014	0.0096
0.014	0.03	0.055	0.03	0.014
0.017	0.055	0.443♣	0.055	0.017
0.014	0.03	0.055	0.03	0.014
0.01	0.014	0.017	0.014	0.01

♣ *Center cell of contact field*

The adoption probabilities of innovation P_{ij} *are defined for the duration of the procedure; they depend on the spatial interaction function chosen to characterize the process.*

c) Probabilities P_{ij} weighted by population N_{ij} in the case where the transmission originates from cell x_i, y_j

	y_1	y_2	y_3	y_4	y_5
x_1	0.056	0.126	0.068	0.168	0.01
x_2	0.014	0.09	0.109	0.06	0.224
x_3	0.134	1.258	1.773	0.383	0.034
x_4	0.0700	0.6321	0.4923	0.3612	0.0840
x_5	0.0192	0.084	0.1848	0.3640	0.0480

d) Probabilities of adoption weighted and standardized in the case where the transmission originates from cell x_i, y_j

	y_1	y_2	y_3	y_4	y_5
x_1	0.008	0.018	0.01	0.025	0.001
x_2	0.002	0.013	0.016	0.009	0.033
x_3	0.02	0.184	0.259	0.0559	0.0049
x_4	0.01	0.092	0.072	0.053	0.012
x_5	0.003	0.012	0.027	0.053	0.007

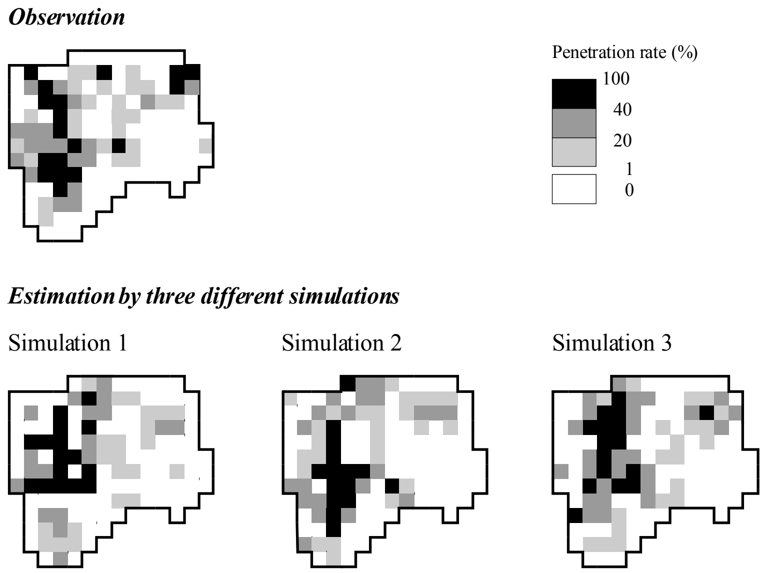
Table 5.2. *Example of contact field with the assumption of a heterogenous distribution of potential adopters*

– the existence of barriers slowing down or blocking diffusion in the zone studied is introduced in the model by considering that when a barrier intervenes in the contact trajectory, the probability that the contact will happen decreases. This decrease is based on hypotheses made on the permeability of the barrier in the case of the process studied;

– the uneven resistance of zones to the innovation adoption is considered by correcting the rule according to which adoption takes place as soon as a message is received. We can, in this case, refer to the propositions later made by Cliff and Ord [CLI 73].

5.3.3. *Simulation procedure*

The simulation of the diffusion process is done by a Monte Carlo type procedure. At each time T_n, the average information field is centered successively on each source cell. A random draw designates the receiving cell. When they can operate at individual level and integrate Monte Carlo type random draw procedures, we can consider that the microsimulation tools could eventually be adapted to simulation approaches such as that recommended by Hägerstrand when he was studying the diffusion of the practice of artificial grassland in Asby county, Sweden (see Figure 5.5).


Observation

Estimation by three different simulations

Simulation 1

Simulation 2

Simulation 3

Figure 5.5. *Simulation of the artificial grassland penetration in Asby county in 1932 (according to [HAG 67])*

Simulation models of spatial diffusion can aspire to predict and anticipate the spatial advances of the process. Several modeling processes especially in the epidemiology field have been developed with this goal in mind. The approach consists of testing the validity of the model first by comparing simulated spatial propagation with observed propagation, then by modifying if needed the model to close the gap between the simulated and observed distributions and, finally, using it to predict the form of spatial diffusion that we may encounter over a period compatible with the hypotheses of the model. Outside this very specific field where the simulation models have not always proven to be as efficient as expected, simulation models applied to spatial diffusion have mostly been used for heuristic purposes. We have mostly been attempting to formalize the knowledge acquired on a process, the spatial diffusion simulation of a phenomenon or a group of simulations, executed with different process development assumptions confronted to an observed diffusion. This is the approach taken by Hägerstrand when he proposed diffusion modeling of artificial grassland in the Asby county in Sweden. In this relatively simple example, the author simultaneously showed the correct convergence of the results of the three successive simulations and the satisfying description of the spatial distribution of agricultural operations using the innovation. The assessment of a certain number of gaps between the observed and simulated distributions sent the author back to a questioning of the limits of the model that was too restrictive according to him, and only integrated spatial proximity properties in the definition of propagation channels. Using the same example, Cliff and Ord [CLI 73] drew slightly different conclusions. They emphasized the spatial autocorrelation of residuals between observed and simulated values and considered that the model had a tendency to overestimate the impact of the diffusion in and around cells that were initially reached, attributing the differences by two causes. The first one, according to them, was due to the fact that the interaction function defining the contact field corresponded to a too slow decrease of the probability of contact between transmitter and receiver over distance. The second gap mentioned by these authors affected the way that the degree of barrier permeability had been defined: impermeable or permeable with a 50% diminishing effect, isolating the territory's peripheries made access to them difficult, thus increasing the chances for underestimating the diffusion.

These investigations have also led to the exploration of possible forms of spatial diffusion, generated by modifying the initial conditions or the process parameters themselves. Placed in this context, diffusion process modeling has evolved into refined interaction functions and simulation procedures that the recent computer modeling improvements should facilitate.

5.4. Analysis models, interpretative models

In 1992, Gould mentioned in epidemiological research that, in spite of simulation tools that are increasingly powerful, predictive modeling, developed to simulate spatial diffusions, did not develop as expected. He blamed this weakness on too much prior ignorance of the processes at work. This comment can be spread to all spatial diffusion simulation studies. For a long time, research mostly focused on the refinement of formalizations and simulation procedures rather than on a detailed and formalized exploration of the processes at work in a propagation movement. Gould invited researchers to use the new information technology tools more efficiently in order to achieve a more systematic and formalized search of spatial forms of the diffusion processes identified and to return to more realistic simulations which are probably more efficient for applied perspectives.

Following these recommendations, spatial diffusion process modeling has been developed in the last few years. We have imperceptibly moved from an "everything happens as if" approach to an approach that will explain how things work, qualifying the mechanisms behind spatial propagation in a better way. This modeling is more focused on either the definition of the spatial form of the diffusion in order to infer hypotheses on the process itself, or modeling of a group of hypotheses that may account for the process studied. These explorations are particularly prolific because they enable us to penetrate the great complexity that spatial diffusion processes show. With the help of two examples we propose to discuss this research direction which should in time greatly broaden our understanding of these processes. First we would like to recall the particular interest in the analysis linked to the location and identification phase of a spatial diffusion process. Then we shall concentrate on the methodology which focuses on modeling the form that the diffusion process takes by using only the spatial or spatio-temporal attributes of the regions concerned or potentially concerned by the process. Thirdly, we will focus on the construction of models whose goal is to interpret a specific spatial diffusion process by using the spatial attributes of places as well as their semantic attributes

5.4.1. *References*

Before any modeling venture, the analysis of the data is often used to identify a spatial diffusion process, to locate its main frames. A few simple questions will generally guide this exploration. Can we locate sequencing of the appearance of a phenomenon that combines spatial and temporal regularity? Has the process started from one or more centers? Has the diffusion been continuous in time?

In his study on the diffusion of mendicant orders in medieval France in terms of traces of urban development, Guerreau [GUE 81] considered the 609 convents created between 1235 and 1450. For each one, the author knows the congregation of affiliation, the founding date and the location. He examines the information obtained at the "department" level, and lists, for each of them, the convents of each of the four main mendicant orders during seven consecutive time periods. A factor analysis done on this chart makes it possible to individualize the stages of diffusion and the spatial models different for northern France and southern France. In the North, a highly hierarchized diffusion corresponds to a quick start and to an early loss of impetus of the largely concentric ring movement from the heart of the Bassin parisian (a quadrangle made up of Sens, Orleans, Chartres and Paris, which were initial movement centers). In the South, on the other hand, no global spatial structure appears. A slower start and more powerful creation flows afterward are associated with several centers located along an axis; Nice, Marseille, Narbonne, Toulouse and La Rochelle. In these cities, the premature creations are relatively dense but there is no discernible spatial diffusion structure surrounding them. From this assessment, Guerreau formulates the hypothesis of a lower interurban integration affected by these 369 southern foundations, and of the existence of several diffusion levels. The author finds a level of cities with a southern urban network relatively well integrated on the one hand, and the level favoring local exchange networks very active but not integrated to the global network on the other hand.

In order to model the bases of the demographic transition development in France in the 19th century, Bonneuil [BON 97] relies on a hierarchical ascending classification of departments, which orders them according to the fertility rate curves registered between 1806 and 1906. With this classification, the author can identify throughout France three centers where the ratio has been continuously the lowest and that were prematurely affected by decline: Normandy and Loire, Champagne and north of Burgundy, and finally the Garonne valley. For each of these centers, the author notes that a rapid decline tendency has occurred internally and by degrees, as in the last center for example where diffusion, starting at Tarn-et-Garonne, would have expanded toward Gironde as well as Haute-Garonne. Rural communities such as Limousin, Brittany, Hautes-Alpes and Corsica were the last departments affected. Based on the assessment of a reduction of the average level and of the dispersion of the departmental fertility rates with the referencing of these regular spatio-temporal tendencies, the author was able to place the first milestones of the research's main hypothesis: the way in which the demographic transition spread over the national territory justified that this propagation was formulated as a spatial diffusion process. On the other hand, this process had to be integrated as it is to the interpretive schema of the transition development in France in the 19th century.

5.4.2. *Models of form*

Beyond these preliminary explorations, modeling can correspond to an explanatory research of the spatial diffusion process rules by the definition of its form. The approach consists of identifying the known functions accounting for the spatial distributions at the different diffusion stages. We generally compare the observed spatial distributions to what they would have been if generated by known formation rules and complying with hypotheses made on the development of the process studied. For example, relating to the electoral diffusion process in western France, Bussi and Mathian [BUS 93] explored the degree of regularity or clustering in location patterns, using the method of nearest neighbors. However, in order to define the form of point pattern coming from a diffusion process, we will not generally favor the reference to a strictly random point pattern (Poisson's distribution). In this way, the binomial distribution will serve as reference when, during propagation, the probability that an element will locate itself in a cell will decrease based on the number of elements already in that cell, the process will generate a more regular scatter than the one from Poisson's law. A reference to a binomial distribution has been tested by Guerreau [GUE 81] in order to find balance points between the pattern of the mendicant convents and the city networks in the north and south of France. The model was found to be relevant only for the north of France (see Table 5.3).

The author demonstrates in fact that, in the south of France, a pattern of convents corresponds in reality to two patterns: one for the convents located in the cities that would best be defined by a binomial distribution and one for the convents linked to small rural towns, related to local monetary networks and situated outside the contemporary urban network. This pattern of small towns corresponds to a strictly random distribution. When the diffusion is accompanied by a tendency for spatial concentration, the negative binomial distribution seems to be more adapted. Often seen with spatial diffusion of certain epidemics, the first location is equiprobable. Then, however, the probability that an element is located in a place depends on the presence of the number of elements already in the place and this probability increases with the number of points already there.

In these examples, temporal autocorrelation is not considered and the distributions are analyzed at each date, regardless of the state of the system in the previous dates.

a) North of France

Years	1235	1250	1265	1280	1300	1350	1450
Number of convents	54	82	107	126	140	165	219
Number of departments with:							
0 convent	16	9	3	2	2	1	0
1 convent	9	8	10	9	5	3	2
2 convents	11	15	14	10	10	9	4
3 convents	2	1	4	8	9	8	5
4 convents	3	4	4	2	5	6	8
...	…	…	…	…	...	...	...
11 convents	0	0	0	0	0	0	2
Correspondence probability between observed and theoretical distribution in the case:							
– of a binomial distribution	0.22	0.02	0.37	0.24	0.94	0.88	0.79
– of a normal distribution	0.22	0.05	0.07	0.40	0.69	0.64	0.95

b) South of France

Years	1235	1250	1265	1280	1300	1350	1450
Number of convents	33	76	129	190	254	324	369
Number of departments with:							
0 convent	26	14	7	4	3	1	0
1 convent	7	9	4	3	2	2	1
2 convents	5	6	10	4	5	4	1
3 convents	4	8	6	10	4	5	3
4 convents	1	2	6	4	4	2	5
…	...	...	...	...	...	...	...
18 convents	0	0	0	0	0	2	3
Correspondence probability between observed and theoretical distribution in the case:							
– of a binomial distribution	0.025	0.025	0.66	0.26	0.013		
– of a normal distribution	0.02	0.61	0.86	0.37	0.58	0.97	0.94

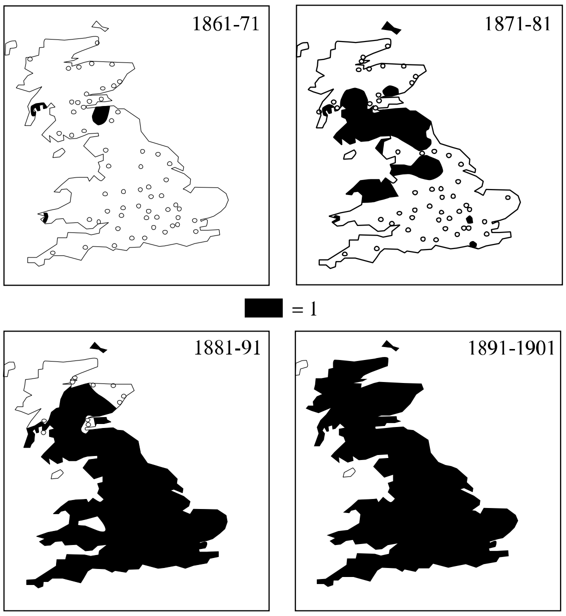
Table 5.3. *The diffusion of mendicant convents in medieval France (according to [GUE 81])*

It is possible to use model families that characterize the process at work by simultaneously taking into account the space and time dimensions. In the large range of possible tools, the use of autoregressive spatial models has been frequent and even more so since the information technology formalizations of the cellular automata type facilitate their development. These models tend to explain the state of a place at time T by taking into account the state of this place and of the surrounding places in the past. Significant parameters of the model of a positive spatial autoregression would reveal diffusion. However, these models that generally account well for spatial changes in time have their own specific issue when they are applied to a spatial diffusion process. In fact, they clearly define the spatial autocorrelation of change; they are not able to discern the characteristics of a diffusion process, i.e. a process that is based on spatial interaction mechanisms. They can compare a growth, to a diffusion that over time adapts increasingly well to a geographical gradient from simple feedback mechanisms able to lead to this type of variation by themselves. Different solutions have been researched to get around this type of problem. As an example, we present the solution imagined by Bocquet-Appel and Jakobi [BOC 97] in a study on spatial diffusion of contraception in Great Britain.

The authors were attempting to define the diffusion channels and the form of the propagation of contraceptive methods by asking themselves if the places that committed at each date to the demographic transition, were distributed randomly, in time and space. From a sample of the counties' administrative centers for which the information was known at different dates between 1861 and 1901, the authors have carried out an almost continuous spatial estimation of fertility all over Great Britain. For each of these spatial distributions and for each date, they have then built an index (Z) which defined the entrance of a place in the demographic transition everywhere (noted by 1 if the entrance occurred and by 0 otherwise). This index is presented in the form of a binary variable indicating if the place has passed (place "in transit") or has not passed, a relative threshold of the fertility ratio variation. In order to observe the diffusion's progression, Bocquet-Appel and Jakobi have considered the cumulative spatial distributions of this new variable Z (see Figure 5.6) for each date.

For a given point, the proximities have been simultaneously defined by two distances, spatial and temporal. The spatio-temporal interaction is tested on this binary variable with the Knox test whose zero hypothesis is the following: are the districts executing the demographic transition in time distributed randomly in space, which would equal the absence of a diffusion process, or are they simultaneously closer on average in time and space, which would reveal a diffusion process.

Cumulative spatial distribution of the cells of the variable Z (0,1): introduction of contraceptive methods for the four periods from 1861-1871 to 1891-1901



Spatio-temporal distribution of the pair of cells which executed the transition for two geographical proximities

Temporal proximity	Geographical proximity			
	Proximity: distance < 383 km		Proximity: distance < 100 km	
	Close cells	Distant cells	Close cells	Distant cells
Close cells	755	391	116	1,030
Distant cells	869	988	111	1,746

Figure 5.6. *Diffusion of contraception in Great Britain between 1861 and 1901 (according to [BOC 97])*

Two distances, one geographical and the other temporal, are fixed in order to respectively determine proximity and, beyond that, separation in each of the two places. In the absence of any information on the geographical radius of the presumed diffusion, two values have been set to define the average geographical proximity between all the pairs of places, 383 km on the one hand and 100 km on the other. Proximity in time has been determined in a way that only the districts executing the transition over the same chronological period can be considered. For a group N of districts, statistic X of Knox shows, among the ($N(N-1)/2$) pairs of districts, those that are simultaneously close in time and space (see Figure 5.6). In general, the method has enabled the authors to individualize a diffusion process, to define its spatio-temporal form by identifying its two original centers and by showing that the process in Great Britain has started in very high fertility zones.

For other researchers, the variables used to simulate the spread of innovation over space do not reduce to spatial or temporal variables. For example, Smith [SMI 04] suggests an event-based process in which the likelihood of the next adopter being in region r is influenced by two factors: the first one being the potential interactions of individuals in r with current adopters in neighboring regions, and the second one being all other attributes of individuals in r that may influence their adoption propensity. The first factor is characterized by a logit model reflecting the likelihood of adoption due to spatial contacts with previous adopters, and the second by a logit model reflecting the likelihood of adoption due to other intrinsic effects. The resulting spatial diffusion process is then assumed to be driven by a probabilistic mixture of the two. These results are applied to a small data set involving the adoption of a new Internet grocery-shopping service by consumers in the Philadelphia metropolitan area.

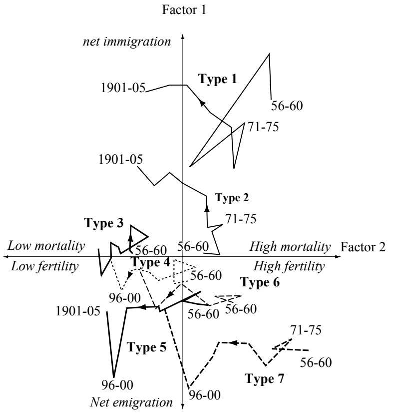
In another way, Grasland and Guérin-Pace [GRA 05] introduced a gravity model to simulate the diffusion of foreign Euro coins during 2003. They estimated the probability density of foreign Euros after one year, but they emphasized a number of difficulties to adjust their model to the distribution observed.

5.4.3. *Explanatory models*

The formalizations mentioned previously show spatial dynamics by mobilizing, in terms of areas potentially affected by the diffusion, spatial properties and even spatio-temporal properties. Another approach consists of trying to interpret a diffusion process as a result of the spatio-temporal combination of several spatial change vectors. In this case, the development of dynamic models combines spatial and semantic variables simultaneously. We will not here go over an overview of explorations. Two specific examples will enable us to simply illustrate the wealth of such research.

The works of Bonneuil [BON 97], relative to the diffusion of the demographic transition on the French territory in the 19th century, are of particular interest in illustrating this type of exploration. The methodology used by the author involves a multivariate autoregressive model which makes it similar to the relatively traditional approaches in econometrics for the analysis of chronological series. Using a hierarchical ascending classification, Bonneuil defines homogenous department classes in terms of their demographic evolution during 10 consecutive five year periods (between 1856 and 1906). The evolution is defined by the variations of fertility ratio, life expectancy and net migration ratio among women aged between 20-24 and 25-29. The author considers that the whole of these types is representative of the space-time of demographic transition in France (see Figure 5.7). He also makes the hypothesis that the changes that occurred in the long term for fertility rates result from a progressive adjustment of these variations to local conditions, themselves resulting from joint effects linked to changes in mortality rates, migration rates, urbanization rates and education levels. By applying this model to the different spatio-temporal types previously defined ("department" classes), the author confirms the strength of the model of a transition space-time and, therefore, of a spatial diffusion of this transition.

a) Types of departmental trajectories



The types of demographical trajectories of the French departments between1856-1860 and 1901-1905 (based on scores of the two leading components)

b) Space-time of the transition, according to the demographic evolution types of the departments

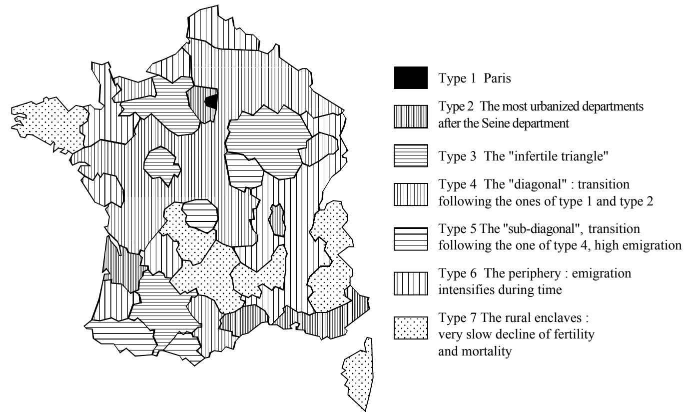


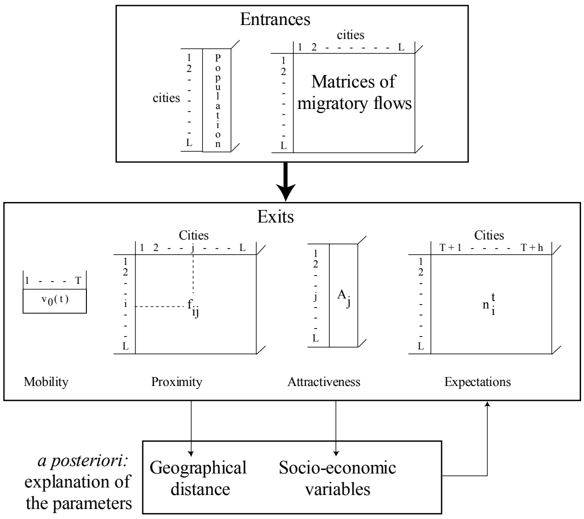
Figure 5.7. *Space-time of the demographic transition in France (according to [BON 97])*

The analysis method of the relations between temporal series has put the emphasis on the joint evolutions of non-stationary temporal series that can be verified in the long term. It enables the author to show that the balance between fertility and evolution of the local environment is done in the long term since short-term interactions only seem to occur in rural parts of the territory, where the importance of the migratory flux toward Paris lessens the influence of local conditions over the long-term variations of fertility rates. The interest in this approach comes from an exploration of temporal adjustments behind a composite spatial diffusion process.

In the tradition of interpretative approaches of diffusion processes, Sanders [SAN 92] explores another way. The author defines and interprets the evolutions during the second half of the 20th century of the attractiveness of French cities. This attraction is seen as an indicator of the diffusion of the economic change in the system of cities. In order to model these attractions, Sanders uses the conceptual frame of synergetics whose objective is to analyze complex systems made up of several subsystems linked together by cooperation relations. The modeling of the dynamics of city populations only considers growth due to migrations internal to the system. One of the originalities of the dynamic model developed is the integration *a priori* of few hypotheses relative to the urban theory and no specifically geographical variable. On the other hand, the calibration of the model provides several specific indicators of the population redistributions between the cities.

Comparing these indicators (proximity effect and attractiveness) with the variables describing the economic and social environment of the system results in the formulation of hypotheses relative to the diffusion process of the economic change of the cities' systems. By using a regression analysis, Sanders tests the hypotheses according to which the attractiveness of cities would depend, in each time sequence, on the importance of their population through the double effect of the combination and saturation mechanisms, and of the preference associated with a city, in terms of its situation and cultural, economic and social characteristics, but independent to its size. The model emphasizes the large inertia of attractiveness which appears to be highly dependent on the hierarchical structure of city sizes and for which the saturation effects have very little impact. The analysis of the residuals of regression, which are the real indicators of the existence of differential attractiveness, clearly shows regional contrasts rapidly evolving in time and that can be compared with spatio-temporal diffusion modes of new economic activities in the system of cities (see Figure 5.8).

Operating the model



Attractiveness decomposition

$$\underbrace{A_i(t)}_{\text{attractiveness}} = \underbrace{Kn_i(t)}_{\text{agglomeration effects induced by the size of the city}} - \underbrace{\sigma n_i(t)^2}_{\text{saturation effect}} + \underbrace{\delta_i(t)}_{\text{residuals: preference}}$$

Diffusion of an innovation cycle

Preferences of the cities during the 1975-1982 period

Change of cycle

Change of preference between 1954-62 and 1975-82

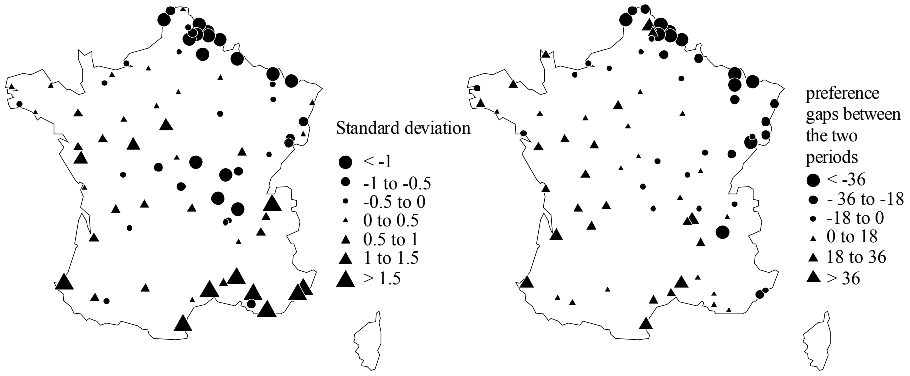


Figure 5.8. *Dynamic model resulting from synergetics concepts (according to [SAN 92])*

The differences of relative attractiveness for the French cities reflect their favorable positions in relation to subsequent innovation cycles. The synergetic model considers these two dimensions. The first dimension is shown through the effect of city sizes and the second is emphasized by the preferential attractiveness indicators. All in all, the concept of synergetics and dynamic modeling of urban populations from interurban migrations have enabled the production of particularly significant indicators (attractiveness and preferential attractiveness) of the spatial diffusion of innovation cycles in the French urban system.

5.5. Conclusion

The state of studies on spatial diffusion modeling reveals an explosion of current research and a certain loss of readability in the analysis. In reality, this trend shows the ambition of these research projects aiming for greatly complex processes. Such an ambition requires that the researchers use new, more adventurous venues because the ones mastered for a long time have probably taught us all they could. So it is less on this loss of immediate readability that we must focus and more on the wealth of these multiple exploratory research projects. Looking forward, this research will give us a completely renewed understanding of spatial diffusion processes.

5.6. Bibliography

[AMA 90] AMAT-ROZE J. M., “Le sida en Afrique Noire: approche géographique d’une infection”, *Revue Belge de Géographie*, no. 4, p. 195-205, 1990.

[ASS 04] ASSELIN S., SEGURA L., “Cartographie animée de la diffusion spatiale, application à la migration des patronymes de la moyenne vallée du Rhône depuis 1891”, *Cybergéo*, 284, 13 p, 2004.

[BOA 40] BOAS F., *Race, Language and Culture*, New York, 1940.

[BOC 97] BOCQUET-APPEL J.P., JAKOBI L., “Diffusion spatiale de la contraception en Grande-Bretagne, à l’origine de la transition”, *Population*, no. 4, p. 977-1004, 1997.

[BON 97] BONNEUIL N., *Transformation of the French Demographic Landscape 1806-1906*, Oxford University Press, 1997.

[BOU 72] BOUDEVILLE J.R., *Aménagement du territoire et polarisation*, M. T. Editions Genin, Paris, 1972.

[BRO 81] BROWN L.A., *Innovation Diffusion, a New Perspective*, Methuen, London, 1981.

[BUS 93] BUSSI M., MATHAIN H., “Etude de la contagion spatiale, formalisation et application”, *Cahiers de Géographiques de Rouen*, no. 38, p. 45-68, 1993.

[CAU 76] CAUVIN R., RIMBERT S., *Lecture numérique des cartes*, University of Fribourg editions, 1976.

[CLI 73] CLIFF A., ORD J.K, *Spatial Autocorrelation*, Pion, London, 1973.

[CLI 81] CLIFF AD, HAGGETT P, ORD J.K., VERSEY G.R., *Spatial Diffusion: An Historical Geography of Epidemics in an Island Community*, Cambridge: Cambridge University Press, 1981.

[CLI 00] CLIFF A.D., HAGGETT P., SMALLMAN-RAYNOR MR. *Island Epidemics,* Oxford: Oxford University Press, p. 563, 2000.

[COH 72] COHEN Y.S., *Diffusion of an Innovation in an Urban System. The Spread of Planned Regional Shopping Centers in the United States 1949-1968*, University of Chicago, Research Papers Series, Chicago, 1972.

[DEG 94] DEGENNE A., FORZE M., *Les réseaux sociaux*, A. Colin, Paris, 1994.

[ELI 97] ELIOT E., "La diffusion du VIH à Bombay", *Espace Population Société*, no. 2-3, p. 253-264, 1997.

[FOL 03] FOLTETE J. C., "Reconstitution d'une diffusion spatiale à partir d'une succession d'états", *L'espace géographique*, no. 2, p.171-183, 2003.

[FRE 85] FREEMAN C., "Long waves of economic development", in *The Information Technology Revolution*, Forester T. (ed.), Blackwell, Oxford, 1985.

[FRI 69] FRIEDMANN J., *General Theory of Polarized Development*, Austin (Texas), 1969.

[GEO 51] GEORGE P., "Critique de la notion de genre de vie", in *Introduction à l'étude géographique de la population du monde*, PUF-INED, Paris, Book 14, 1951.

[GOU 68] GOUROU P., *Pour une géographie humaine*, Flammarion, Paris, 1968.

[GOU 92] GOULD P., "Epidémiologie et maladie", in *Encyclopédie de la Géographie*, Bailly A., Ferras R., Pumain D. (eds.), Economica, Paris, p. 969-990, 1992.

[GUE 81] GUERREAU A., "Analyse factorielle et analyses statistiques classiques: le cas des ordres mendiants dans la France médiévale", *Annales – Economie Sociétés Civilisations*, no. 5, p. 869-912, 1981.

[GRA 05] GRASLAND Cl., GUERIN-PACE Fr., "La diffusion spatiale, sociale et temporelle des pièces euros étrangères: un problème complexe", *Actes des Journées de Méthodologie Statistique*, Paris INSEE, 2005.

[HAG 52] HÄGERSTRAND T., "The propagation of innovation waves", *Lund Studies in Geography*, B series, no. 4, 1952.

[HAG 67] HÄGERSTRAND T., *Innovation Diffusion as a Spatial Process*, Translation of Pred, Chicago University Press, 1967 (E.O. 1953).

[HAGG 77] HAGGET P., CLIFF A.D., FREY A., *Locational Models,* Arnold, Bristol, 1977.

[HAI 82] HAINING R., "Interaction models and spatial diffusion processes", *Geographical Analysis*, no. 2, p. 95-108, 1982.

[HAI 83] HAINING R., "Spatial and spatial temporal interactions models and the analysis of patterns of diffusion", *Transactions Institute of British Geographers, New Series*, no. 2, p. 158-186, 1983.

[HAL 88] HALL P., PRESTON P., *The Carrier Wave New Information Technology and the Geography Innovation 1846-2003*, Sage Publishers, 1988.

[HUD 69] HUDSON J.C., "Diffusion in a central place system", *Geographical Analysis*, no. 1, p. 45-58, 1969.

[KIM 05] KIM D.S., CHUNG H.W., "Spatial diffusion modeling of new residential area for land-use planning of rural villages", *Journal of Urban Planning and Development*, no. 3, p. 181-194, 2005.

[LAK 84] LAKE L.A., SEYDOU-NOUROU-TOURE E.M., *L'expansion du bassin arachidier au Sénégal, 1954-1979*, IFAN, Dakar, 1984.

[LAN 83] LANGDALE J., "Competition in the United States long distance telecommunications industry", *Regional Studies*, p. 393-409, 1983.

[LEV 58] LEVI-STRAUSS C., *Anthropologie structurale*, Plon, Paris, 1958.

[LYT 89] LYTT I., "Spatial diffusion of the human immunodeficiency virus infection epidemic in the United States, 1985-1987", *Association of American Geographers*, no. 1, p. 25-43, 1989.

[MAH 85] MAHAJAN V., PETERSON R.A., *Models for Innovation Diffusion*, Sage, Newbury Park, 1985.

[MEN 70] MENDRAS H., *La fin des paysans*, A. Colin, Paris, 1970.

[MEN 79] MENSCH G., *Stalemate in Technology*, M.A. Ballinger, Cambridge, 1979.

[MOR 65] MORRILL R.L., "The negro ghetto, problems and alternatives", *Geographical Review*, no. 55, p. 339-361, 1965.

[MOR 89] MORRILL R., GAILE G.L., THRALL G.I., *Spatial Diffusion*, Sage Publications, Scientific Geography series no. 10, London, 1989.

[MUR 77] MURRAY G.D., CLIFF A.D., "A stochastic models for measles epidemics in a multi-region setting", *Transactions Institute of British Geographers, New Series*, no. 2, p. 158-174, 1977.

[PAL 91] PALMA DE A., DROESBEKE J.-J., LEFEVRE C., *Modèles de diffusion en marketing*, PUF, Paris, 1991.

[PED 70] PEDERSON P.O., "Innovation diffusion within and between national urban systems", *Geographical Analysis*, no. 3, p. 203-254, 1970.

[PER 57] PERROUX F., *Théorie générale du progrès économique*, ISEA notebooks, series 1, 1957.

[PRE 66] PRED A., *The Spatial Dynamics of US, Urban Industrial Growth 1800-1914. Interpretive and Theoretical Essays*, MIT Press, Cambridge, 1966.

[PRE 73] PRED A., TORNQVIST G., *Systems of cities and information flows*, Gleerup, Lund, 1973.

[PRE 75] PRED A., "Diffusion organizational spatial structure and city system of development", *Economic Geography*, no. 51, p. 242-268, 1975.

[PUM 97] PUMAIN D., SAINT-JULIEN T., *L'analyse spatiale, Volume 1: Localisations*, A. Colin, Cursus Collection, Paris, 1997.

[PUM 01] PUMAIN D., SAINT-JULIEN T., *Les interactions spatiales, Volume 2: Interactions*, A. Colin, Cursus Collection, Paris, 2001.

[RAV 98] RAVENEL L., *La géographie du football de haut niveau en France*, PUF, Paris, 1998.

[ROG 83] ROGERS E.M., *Diffusion of Innovations*, The Free Press, New York, 1983 (3rd ed.).

[SAI 82] SAINT-JULIEN Th., *Croissance industrielle et système urbain*, Economica, Paris, 1982.

[SAI 85] SAINT-JULIEN Th., *La diffusion spatiale des innovations*, Reclus Mode d'Emploi, Montpellier, 1985.

[SAI 92] SAINT-JULIEN T., "Diffusion spatiale", in *Encyclopédie de la Géographie*, Bailly A., Ferras R., Pumain D. (eds.), Economica, Paris, p. 577-598, 1992.

[SAI 99] SAINT-JULIEN T., *Services et commerce. Atlas de France, volume 11*, Reclus-La Documentation Française, Montpellier-Paris, 1999.

[SAN 92] SANDERS L., *Système de villes et synergétique*, Economica-Anthropos, Paris, 1992.

[SCH 34] SCHUMPETER J., *Theorie der wirtschaftlichen Entwicklung*, 1934 (1st French translation, 1935, *Théorie de l'évolution économique*).

[SCH 35] SCHUMPETER J., *Business Cycles: A Theoretical Historical and Statistical Analysis of the Capitalist Process*, New York, 1935.

[SMI 04] SMITH T.E., SONG S., "A spatial mixture model of innovation diffusion", *Geographical Analysis*, no. 2, p. 119-145, 2004.

[SOR 48] SORRE M., "La notion de genre de vie et sa valeur actuelle", *Annales de Géographie*, no. 306, p. 98-108, no. 307, p. 193-204, 1948.

[STR 98] STRONG D., SOULE S.A., "Diffusions in organizations", *Annuals Review Sociology*, vol. 24, p. 265-290, 1998.

[THO 75] THOUVENOT C., *Les consommations et les habitudes alimentaires dans la France du Nord-Est*, RT, Lille, 1975.

[TOR 70] TÖRNQVIST G., *Contact Systems and Regional Development*, Gleerup, Lund, 1970.

[VAN 81] VAN DER LEEUW S.E., "Information flows, flow structures and the explanation of change in human institutions", in *Archeological Approaches to the Studies of Complexity*, Instituut voor Prae-en Protohistorie, Amsterdam, 1981.

[VAN 89] VAN DER LEEUW S.E., TORRENCE R., *What's New? A Closer Look at the Process of Innovation*, Unwin and Hyman, London, 1989.

[VID 11] VIDAL DE LA BLACHE, "Les genres de vie dans la géographie humaine", *Annales de Géographie*, no. 111, p. 193-212, no. 112, p. 199-304, 1911.

[VIL 04] VILALTA Y., PERDOMO C.J., “The local context and the spatial diffusion of multiparty competition in Urban Mexico 1994-2000”, *Political Geography*, no. 4, p. 403-423, 2004.

[WEB 78] WEBBER M.J., JOSEPH A.E., “Spatial diffusion processes 1: a model and an approximation method”, *Environment and Planning*, no. 6, p. 651-666, 1978.

[WEB 79] WEBBER M.J., JOSEPH A.E., “Spatial diffusion processes 2: numerical analysis”, *Environment and Planning*, no. 3, p. 335-348, 1979.

Chapter 6

Spatial Microsimulation Models

6.1. Introduction

The first reference to a microsimulation model in social sciences goes back to studies from the American economist Orcutt and his team in the 1960s [ORC 57, ORC 61]. His objective was to provide national forecasts on the evolution of employment in different economic sectors. Disapproving the traditional macroeconomic models, he developed a model at the household level, then aggregating the results in order to produce forecasts at an aggregate level. The necessity of considering basic entities in order to understand the form and evolution of observable organizations and aggregates at a higher level is an old subject of debate in several disciplines such as ecology, biology, economy and sociology. The theoretical debate is old but the increase of computing power has given it a new perspective. In addition, with artificial intelligence methods, it is possible to formalize the behavior of "agents" that have autonomy, adaptation and interaction capabilities [FER 95] and these new perspectives have led to the development of the use of simulation in all social sciences [BAL 00, GIL 99].

A general label for this kind of models is "individual-based". Depending on the disciplines and applications, the term can mean "individuals" of extremely diverse nature (cells, particles, living beings, countries or planets). On the other hand, the term microsimulation mostly refers to basic entities that make sense in economy, demography and sociology, such as the individual, household, family and firm. It is also widely used in the transportation and health sectors.

Chapter written by Einar HOLM and Lena SANDERS.

Most microsimulation models have been applied and developed for questions without any specific focus on the spatial dimension of the modeled phenomenon. Usually, two levels are involved in most social science applications: the micro-objects level (individuals, households or firms) and some aggregates of the same micro-objects for the geographical region or country studied. Microsimulation is often used to make forecasts and/or test the impact of a changed policy for the evolution of a given phenomenon globally and individually. Evaluating the effect in *k* years of changed taxes or public expenses for retirement or family allowances on the consumption and lifestyle of different types of households might be an example.

In the field of spatial analysis, one often explicitly focuses on the intermediate, meso-geographical observation levels in between the micro-objects and the global level, and the distribution of the phenomenon studied is described based on such spatial entities. If the objective is to simulate the evolution of a population's spatial distribution, the modeler must choose between a representation at the level of the chosen spatial aggregates (cities, municipalities or districts) or at the basic entity level (individuals, jobs, homes). Each of these choices implies certain hypotheses about the organization and functioning of the studied system that are not always explicitly stated in the applications and needs to be deliberately contemplated and motivated.

The identification of the most adequate level of observation, representation and modeling is not always immediately self-evident and does not necessarily only depend on the question raised. Before going through the basic principles of a microsimulation model (see section 6.3) and presenting concrete applications (see sections 6.4 and 6.5), the choices of objects and modeling levels, and their consequences are discussed. The notions of basic entity or "micro-object" and the existence of "atoms" that would emerge as obvious micro-objects in different systems are first discussed. The respective advantages of an "individual-based" approach and of a meso-geographical model are then evaluated. The specific advantages of formulating the model from basic entities are compared to the alternative of basing the analysis on some aggregation of these basic entities.

6.2. Choosing the aggregation level for modeling

Choosing a level for modeling means defining the micro-objects where the mechanisms of change can be identified and understood. The discussion refers on the one hand to conceptual questions (section 6.2.1) and on the other hand to the methodological framework in which the model is developed, as well from a theoretical (section 6.2.2), thematic (section 6.2.3) or technical (section 6.2.4) perspective.

6.2.1. *"Micro-objects" and spatial analysis*

In a social system, the individual is often seen as the basic entity, the fundamental atom. A small logical aggregate corresponds to the household. However, the household can also be considered as a micro-object which is indivisible and the most adequate for the problem under study. It is the same way with other, more important, aggregates such as a firm, a school or even a city. In order to determine if such entities can be considered as micro-objects, we must answer the following question: is what we are studying an attribute of the whole object or is it the result of the sum of the properties of lower level components of the object studied? By applying this "subsidiarity principle", each system should be divided in lower level components as long as the behavior, process or characteristic studied (or its main drivers) varies between the components.

6.2.1.1. *Arguments for choosing a modeling level*

Most human activities are strictly bound to be performed by all parts of the physical body simultaneously. If you move to another place, you either do it with the entire body or not at all, there is nothing in between. It is not an option to stay with the head and leave with the legs. The same goes for most human properties. If a person gains a new education or a new profession or a new political view, it is pointless to claim that only the left side of the body has this new education, etc., because if so the right side inevitably has to stick with all aspects of the properties of the left side. Neither the person nor anybody else can make any such distinction between the two sides – the property is a property of the entire person and nothing else. If you break your left arm, however, it makes some sense to say that it is your right side only that open doors, etc. Different parts of the body might sometimes be regarded as micro units by themselves, but the freedom of action for those parts are quite limited. The right arm cannot leave the rest of the body and open a door 2 m away by itself and, more importantly, it cannot open the door anyway unless you, as a complete autonomous person, order it to do so.

Walking down the line from human, via biological towards pure physical properties of the body, like its mass, widens the separation perspective. The mass measure and impact (i.e. on acceleration) is exactly the same whether or not each molecule of the body is still in a living human or if they have dissolved into a fluid in a container. This property (mass) of the living organism is nothing more than an aggregation of the same property of each molecule (micro unit) in the body. Nevertheless, from an anthropogenic perspective (which is the only perspective of interest here), the individual human is a very tight and distinct system, almost all properties and activities of a human that are of interest to model are properties of the entire person. It could even be argued that it is the system of organized connections between cells and molecules in the living organism, as created by evolution, that is the human individual regardless of whether or not the micro units are based on carbon chemistry or something else (with different masses, etc.). This is why individual humans are the preferred micro unit in microsimulation (and in social science in general, as well as in humanities and politics etc.).

Walking the line "upstream" from the single individual towards organizations with several individuals reveals potential micro units that are not as tightly organized internally as the human body. Such organizations are often defined by relational properties or by common properties. A relational property connects objects (mainly individuals) temporarily or for longer periods. The property labeled "mother" connects an individual with a certain other individual and so does the property labeled "partner". "Mother" is a directed relation (your mother's mother is a third person) while "partner" in most cultures signifies a symmetric exclusive relation; you yourself and nobody else is your partner's partner. Other relations connect individual persons to other individuals like son, daughter, friend, relative, or to other object classes (aggregates of persons or something else) like family, school, place of work, neighborhood, city, region, country of residence, etc. In most cases, you belong to just one family, one school, etc., but the family or school might also contain other individuals. The person belongs to one family, living in one neighborhood in one city, in one region in one country. As soon as several relational types are considered jointly, the strict hierarchy is inevitably broken. The members of a family might belong to different schools and places of work in different neighborhoods in different regions, etc.

However, the fundamental distinction is the one between all relations as above and "solitary" attributes of the individual like age, sex, income, education, profession, etc. They are strictly properties of the individual only and do not contain any relational information. They can be aggregated by any dimension (space, gender, etc.) without other problems than statistical ones. On the individual level it is possible to define solitary aspects of the relational properties. A person is married to a specific other person but by that is also "married". The person works at a specific place of work but is by that is also "working", etc. The relational property contains much more information compared to the solitary aspect, not least since all properties of the person/object pointed at by the relation becomes available as attributes of the first person. Partial impacts like that of parents' education on children's education can easily be found. As soon as the individual properties are aggregated into tables with few dimensions or maps (districts), almost all of that extra relational information is lost; a table can almost only carry aggregates of solitary attributes.

In some cases, the basic decision making entity comes "naturally". If we are attempting to model the individual factors intervening in the choice of transport mode for the home to work trips, then the commuters constitute the obvious basic entities to consider. The aggregation of these choices gives a good indication of the demand in terms of transport. If the object of study is the housing market, houses as the supply units and individuals as demand units appear "naturally" as the most logical basic entities. In the same way, for a study of distribution of production volume between firms, it is "natural" to consider the firm as an atom, the basic entity. In fact, many important driving attributes such as the branch, the value of the product, the profit, etc., will rather characterize the firm globally and not a specific component like a particular worker. Similarly, certain spatial entities can emerge as basic entities. The "behaviors" of a district, a town or a country can partly depend on agencies based only and strictly on these levels and not on lower level actors. For example, to determine the location of a new airport or choose which country will host the Olympics, the decision is clearly made on the macroscopic level, although also basically performed by individuals delegated to represent the decision body. When that is the case, it might be appropriate to use the district, the town or the country as one of the basic modeling entities.

So, in most cases, decision making is based on mechanisms operating simultaneously at different levels. When the objective is to identify and represent the reasons for a migration event as well as to model that activity (example 1), some driving explanations are to be found among the characteristics of the individual (gender, age, education), as well as in his relations with other individuals (accounting for the spouse's wishes, for example), and in the attributes of his home (dimension, quality) or his neighborhood (services, accessibility). The characteristics of the family, the home, the neighborhood and local culture are all attributes characterizing the environment of the individual. Such attributes might very well be represented as properties of separate instances of other higher level units than the individuals in a simulation. If the values of such relational environmental attributes are common to many individuals, they will probably interact with the individual properties of the person himself and therefore the impact can be different for different individuals. The effect of distance from the potential new house to the person's job is probably different if the person owns a car or not. This kind of individualized causality can easily be accounted for with an individual representation. The same argument holds for all levels of decision making.

Indeed, if the goal is to determine in which municipality an equipment should be located (example 2), the characteristics of each municipality, will be examined, like density and distance from any other object influencing the outcome (city, other equipment, highway interchange). Other potential forces are more pure aggregates of individual properties for inhabitants in the municipality like average household income, net migration, farmers' share, etc. Thus, several different attributes characterizing the neighborhood and the municipality's geographical context will be considered.

In each of these two examples, the individual and some higher level organizations are involved. The higher levels appear in two variants: I. as a pure container for indicators entirely based on aggregates of individuals; and II. as intrinsic properties of the level/organization. In the first example, the choice of the individual as a basic entity of the modeling will make more sense if the *combination* of all the higher level and the individual attributes for a specific individual is unique and by that has a specific influence on the individual's decision to migrate. On the other hand, if there are large regularities in the migratory behavior within certain categories of the population, these categories might be considered as the relevant micro-objects. That decision can also be the result of practical necessity. If no individual migration observables are available, then there is no other option. If individual data is at hand, then there is no point in not using them also in cases when most drivers and consequences operate from and to higher level organizations. It is unlikely that the individual variation in attribute values within the groups does not discriminate any part of migration behavior (see discussion below). In the second example, the municipality level can be considered as long as the most discriminating factor refer to the attributes (profile, accessibility) of the municipality, again especially if no information on finer or individual level is given. If, instead, the diversity of individual properties by themselves discriminates behavior, then, for understanding the location of new equipment, the aggregate level of the municipality would not be enough for representing the question in an explanatory simulation model.

6.2.1.2. *Individuals as the favored micro-objects in spatial microsimulation*

Returning to the different organizations above the individual and their eventual role as micro units, when is a family, for example, the suitable micro unit of analysis and simulation? For what purpose can it be regarded as a tight organism acting as a single entity with joint behavior and aspirations like the individual body with its cooperating cells? Location is a relational property that by definition is the same for the family and all its individual members. Is that enough? The family can be given solitary attributes based on its members like size, type, dwelling, total income and consumption, etc. Co-location is almost the definition of a family or at least of a household. So, for a cross-sectional description of settlement and income distribution, families might work as suitable micro units.

However, simulation is about change, not about one pattern. Change of location of the family is more often than not simultaneous with formation and dissolution. It is not the family that decides whether a young person moves from home, it is the person him or herself. It is probably rare that a divorce is a family decision – it is rather about one adult wanting and deciding to move from the family. Having a baby might be a family decision but normally is not and then might rather be the event that triggers family formation in a situation when it is not yet created; moving together has to be a joint decision by two individual persons, but that is before the family is created and exists. Decisions about education, work, etc., are of course primarily individual while influenced by family member properties.

So, probably most family-related events and properties one would model are initiated from individuals rather than from the family as such. If instead the individual person is maintained as a micro unit, the family then becomes just another relational property of the individual. Sometimes that property will change based on individual decisions influenced among other things by the set of properties and decisions of each other's family members.

Most microsimulation models defines their micro units based on decision making which imply searching for the level where the basic decisions about the studied phenomenon are made – on what level is the core "decision making unit" to be found? As discussed above, more often than not, this seems to be on the individual level.

Walking further "upstream" in terms of number of individual members gives candidate decision making units like a firm, a school, a club, a city, a region, a national state, etc. Formally, a firm is a hierarchical organization with all decision making power in the hands of the owner. The whole point with a firm as an organizational entity is to avoid the heavy load of transaction costs involved in having the production performed on micro markets cleared after negotiations by self-employed individuals for each sub-task in the internal production chain. Instead, all of that is replaced by an efficient internal planned economy entirely decided upon by the owner as delegated to his CEO. Therefore, it would be easy to regard the firm as a tightly organized system of employees and equipment – a decision making unit on its own level.

However, the owner and CEO are not dictators and the performance of the firm is highly contingent on conditions they cannot control entirely. One of the most important is quality of, and motivation in, their labor force. This requires a certain degree of freedom and power to decide among the employed, and so again, as the member individuals, the workers, makes decisions influencing the firm, they are also partial decision making units with respect to firm performance.

Secondly, larger firms often have several working places in different locations. Not only does the spatial configuration influence performance but it is also the case that the working places are focus points for the firm's most important resource: the labor force. If the working places are separated by large distances, they are, from the point of view of the employed, the important part of the firm. As soon as spatial conditions and outcome are important objectives for the simulation, it is quite reasonable to instead use the working place as a micro unit, despite most decision making power belonging to the firm and its owner.

Thirdly, the CEO is also nothing but a human. In analogy with the family as discussed above, the firm could be represented as a property of each individual owner, albeit with very different abilities and powers. If the simulation goes deeply into the decision making process it becomes quite strange to apply decision theory just to the abstract firm entity when all the time it is individual humans in management and elsewhere that have aspirations and knowledge and are actually always the ones that individually take all decisions on behalf of the firm.

Modified versions of the family and firm discussion above can be applied on the other social organizations as well: on schools, clubs, cities, regions and national states. This would probably reveal that individuals should often be the basic micro unit based on either their own large influence on the organization's behavior or at least that they have a significant influence on their own relation to the organization, contrary to the cell in the organism.

But is that the full story? If so, it would give full credit to Weber's view in the century-long individualist contrastructuralist discussion on society. Weber and Durkheim have come to embody opposite views in that debate. Durkheim argues that society is something entirely different from its individuals and that its properties cannot be explained by reference to the properties of the individuals, while Weber is often cited as saying there is no society beyond the aggregate of its individuals [BOM 04].

Even if it were the case that individual decisions dominate on all levels and that the specific, remaining decision power of the family, firm, city, etc., is smaller than the aggregate of their members' individual decision power, a blunt majority principle would entirely remove that part of the influence. A pure individualist society without any permanent social organizations except perhaps day-to-day micro markets between neighbors for immediately available commodities would be the consequence of not giving up any individual influence towards collective organizations. This is at best a society of hunters and gatherers, and that is not the society we aim to simulate. On the contrary, the target society is the contemporary society with all its fragile relations constituting historically emerged social organizations requiring the individual to release decision power to the collective. So, although the higher level relations based on power individuals have given up, are relatively weak and are not as tight as the relations between cells in the body, together they create the difference between a hunter and gatherer society and the present one. That is a long way from Weber towards Durkheim, but not at all entirely.

The obvious conclusion, as exemplified earlier, is that there is a division of power between the levels including the individual and preferably the influence from all relevant levels/organizations should be accounted for while modeling both individual and societal change. In most cases, that division gives the individuals at least a certain influence on aggregate results. The choice of a smallest micro unit above the individual level implies that such ”individual deviations” are regarded as irrelevant for the simulation task, hence the argument above for a “subsidiary principle” in relation to the levels.

The division of power between a large set of social organizations with different members influencing individual life and societal outcome gives another entirely practical reason for maintaining the individual as a micro unit, as the smallest object type in the simulation. Only the individual is a common denominator to all the different organizations he or she belongs to. The individual functions as a convenient integration key between them all, as well as in cases when it is appropriate to assume that the bulk of influence stems from different higher level organizations. Then, each person carries relational properties that indicate belonging to certain families, schools, places of work, cities, regions, etc., giving immediate access to the set of properties of each organization on each level relevant for the particular individual. In no way does choice of the individual as the smallest micro unit exclude representations of higher level organizations as single objects with their own properties and behavior in the same model; objects that partly develop based on intrinsic, non-individual properties and actions.

So far, the discussion of an appropriate level for micro units in microsimulation has not considered contemporary research obstacles. While it is easy to claim that the individual is the best and most suitable smallest micro unit from a theoretical, conceptual and practical point of view, it is simply not an option to base an individual representation on located observables in most countries. Large or full scale individual, longitudinal register samples or census data for research are only available in very few countries, such as Sweden. Many countries instead have made special, often survey-based random individual samples available for research with and without microsimulation, but they are often too small to enable any resolution of spatial variation in drivers and outcome. For a recent overview, see [GUP 07]. So, for most countries, the only data available for spatial microsimulation are disaggregated two or three-dimensional tables with at best one dimension representing regions, effectively removing all relational and most solitary information from the underlying secret individual data records. A surrogate solution is to construct a synthetic sample based on and consistent with all tables. This does not recreate the lost information but it gives a platform for experiments where it is important to at least maintain individual heterogenity and theoretical and/or survey-based assumptions on individual behavior responses.

Since this is the most common situation, earlier also in Sweden, it has of course influenced minds in the formation of concepts and theory. The hope is that the smallest thing that can be observed is also the relevant micro level for representation, causality and modeling. An additional force in the same direction is that one is almost never interested in individual outcome, results are not results until aggregated, as averages by region, by region and age, etc., generally like the information given in the supplied aggregate data tables. It would be convenient if the model could be constructed on the same level of aggregation as are the aspired results.

It is easy to demonstrate that microsimulation enables detailed results regarding distributions over demographic and spatial dimensions that are not possible with aggregate models. However, it is still a hurdle to effectively demonstrate that macro results also become better and less biased if derived as an aggregation of micro process outcomes, as is the case in microsimulation. Is it necessarily the case that calculations of the development of income per capita, unemployment level, segregation level, the proportion of elderly or gross fertility in a country replicates observed levels better if modeled with microsimulation compared to macroeconomic or macrodemographic models? The individual variation filtered away in a macro representation is sometimes close to random and is not heavily correlated with anything else; if so, the micro representation adds little to the aggregate, average outcome. The current state of bad data sources, estimation shortcomings and specification errors might easily create heavily biased microsimulations with an error component larger than any systematic aggregated impact difference. This gives feed to the disaggregated modeling opposition that is hard to dispute.

Implicitly, such observations might also partly explain a common conceptual confusion between aggregate and structural conditions. The aggregate of many individuals in one or two dimensions is used as an indicator of a structural condition beyond the apparent control of an individual agent, like income (and power) distribution between individuals and other agencies in a country. Then, the idea is that “power structure” in the spirit of Durkheim is something entirely separate from individual action while heavily influencing individual lives. However, the eventual existence of such structures is not necessarily revealed by the aggregate data observation. Nothing in that excludes at least a partial impact from individual action. There certainly exists early innocent data free agent-based modeling (ABM) applications claiming creation of the emergence of a social order from nothing but randomly assigned individuals with location and two or so aspirations as their properties. This has given additional arguments for those claiming that ABM and other micro models ”strongly tend towards an individualist view of the social world”, i.e. on a bias towards methodological individualism, overestimating the power of the individuals in societal development (see, for example, [OSU 00]).

These few examples show that the choice of the basic objects for an analysis is not always obvious *a priori*, even when question and context are given. It is therefore important to think about the respective roles of the chosen theoretical frame and of the hypotheses on the drivers of change and the technical constraints related to the support of the chosen modeling.

6.2.2. *Theoretical point of view: interactions and emergence phenomena*

A fundamental hypothesis based on the paradigm of complex systems is that the play of *interactions* between lower level entities during evolution eventually results in the *emergence* and durability of new, more complex forms of interaction, visible as organizations at a higher level of observation. This theoretical frame, linking the different observation levels of a phenomenon, can also be applied at different geographical scales, if the lower level is made up of spatial entities or if it is made up of individuals as human beings.

In the case of spatial entities (towns or cities, for example), the interactions refer to flows (of people, of products) and exchanges (of information, of services). These *spatial interactions* are the result of the areas' respective resources and possibilities, and the relative locations of these areas in relation to each other. They contribute to the generation of forms like center-peripheral gradients, discontinuities, concentrations, or at another level, hierarchies and networks.

If, on the other hand, the basic entities are individuals or households, the interactions operate on an individual level. They consist of relations of influence, cooperation and avoidance, and result in events (marriage, birth of a child) and interdependence relations between individuals (the migration of a person who is a member of a family most often results in the migration of their spouse and children). The interaction could also be indirect (the departure of a family means the vacancy of a home constituting an opportunity of moving in for another family), or concern interdependences between the decisions of a person and her own past trajectory (date of access to property influences the decision to migrate, the birth date of previous children influences the decision of having another child, etc.). The result of these interactions can lead to a profile change of a residential neighborhood or the appearance or new patterns of home to work trips and thus contribute to the emergence of spatial organizations, structures, which have a certain meso- and macro-geographical logic. The combination of some behaviors can be sufficient to initiate a change that will be visible in the long term at a meso-geographical level. For example, the installation of a few particular households in a district can make this area attractive (or on the contrary repulsive) for a different category of the population than the traditional population and start a feedback loop that will translate over time in a change in the district profile. Imitation and segregation mechanisms at the individual level can thus result in the emergence and reinforcement of new structures at an aggregate level (such mechanisms are at the base of certain ghettoization processes).

New spatial structures can be interpreted as emerging from behaviors, choices and interactions traceable at an individual level, or at the level of basic geographical entities. The choice of the theoretical frame of complex systems to model evolution of geographical space does not imply *a priori* a preferred observation level. This choice implies the consideration of *several* levels in the modeling but does not suggest anything on the definition of these levels.

6.2.3. *Thematic point of view: the driving role of the inter-individual diversity*

A certain number of studies are based on the hypothesis that aggregated information is sufficient to understand and identify the causes for change. There are then several dynamic models formalized at a meso-geographical level that are based on the hypothesis that spatial interactions play a driving role in the dynamics of the distribution of human populations. These models have been developed with various formalisms: differential equations [ALL 97, CAM 86, PUM 89], cellular automata [CLA 97, ROY 96, WHI 97], multi-agent systems [BUR 96, SAN 97], etc. However, all are based on the same basic hypothesis that there are such regularities at a meso-geographical level that the characteristics of the areas, whether they are inherent to this level of observation or the result of an aggregation operation, are sufficient for the understanding of the evolution of the studied phenomena. If the elementary decisions are actually taken by individuals, the hypothesis is that it is not necessary to understand each of these decisions and that it is enough to focus on their resultant. In that case, it is presumed that the diversity of the individual responses to a same context will result in compensation effects and therefore without repercussions on the higher level structures.

Other research underlines that this *diversity* of individual behaviors plays an important role in change and therefore formalizes the model at an individual level. A simple example will illustrate this approach: let us presume that in order to model the evolution of the population at district level over the next decade, we develop a model based on the average age of the residents of each district. Let us suppose that the average age is 30 for districts A and B, which also have the same number of residents. They will therefore have similar estimated values for the evolution of their population. If, within this same average, we have a relatively homogenous population in terms of age in one town and in the other an overrepresentation of young children and old people, the observed evolutions may be very different. The example is simplistic but the logic is the same in numerous cases. When only one dimension is considered, as in this simple case, it is easy to disaggregrate the model relatively to this dimension, in other words to consider the distribution of the population by age groups in the case of this example. Multidimensional diversity is more difficult to manage conceptually and technically without individual representation.

The work of [AXT 02] gives a convincing example of the driving role of inter-individual diversity. Aiming at simulating the evolution of the Anasazi population from 200 to 1,300 in Long House Valley and the abandonment of the Valley at that latter period, the authors developed an agent-based model. The micro-objects in this model were the households. As long as no inter-individual diversity was introduced in terms of demographic and nutritional behavior, the simulated evolution of the total population did not correspond to the "observed" one (as estimated by the archeologists). There is, however, one important difference between that kind of agent-based modeling and microsimulation. While cellular automata and most agent-based models disregard individual heterogenity in terms of empirical data, they are still formulated on an individual level and test hypothesizes regarding individual interaction. Individual heterogenity is maintained but the distribution between individuals is assumed to be the same for all.

The correct level for the development of a model then depends on the hypotheses that we make *a priori* on the role of diversity at the different geographical levels considered. It is important to identify the "driving diversities", those which are at the roots of change. They correspond to the relevant level to consider, not the level associated with the results of the study.

6.2.4. *Technical point of view: management of information tables*

The discussion also includes a technical aspect. *A priori*, a model developed at a meso-geographical level is more parsimonious in data requirement than a microsimulation model. However, if it is necessary to consider a finer spatial resolution and to take into account the interactions between the considered entities and a certain level of disaggregation of the attributes, certain meso-geographical disaggregated[1] models can lead to gigantic information tables. Let us consider N_{ij}^{mu}, the general term of a flow matrix, representing the number of people of social category *m*, of age group *u* residing in *i* at date *t* and in *j* at date *t* + 1. This matrix is relatively easy to manage if, for example in France, we consider entities as regions or cities and if the number of age groups and social categories considered are not too high. If we consider the 95 departments, a dozen age categories and a dozen

---

[1] The term "disaggregated" is used here when we consider a population according to different categories, possibly defined at a very fine level (individuals of such age class, of such education level and of such social category, for example) but without going down to the individual level.

social categories, the matrix reaches a million elements. If we consider all 36,000 French municipalities, the number of elements in the matrix (several billion elements to manage, even if the majority is zero) is much greater than the French population, even by multiplying it by the number of attributes considered here such as age and social category and the municipality of residence. The consideration of a new attribute would result in an additive increase of the number of elements in a microsimulation approach and in a multiplicative increase in a disaggregated model. As a result of this, a variable such as age can be considered in its continuity, when it is considered as an attribute of an individual and in a less precise way when we are considering all the individuals of one age class.

In conclusion, when the number of attributes to consider is higher, the microsimulation approach technically presents relative assets (see Box 6.1). Faced with a given problem, it is necessary to determine if a global perspective is sufficient, if it must be enhanced from certain disaggregation and when the substitution to a microsimulation is more relevant than a too advanced disaggregation. By applying all the models to one case, we would be able to evaluate the extent of successive improvements of modeled dynamics and to compare a meso-geographical disaggregated model with a microsimulation, in terms of time of development, calibration and outcome quality. Such a systematic comparison has not been published in human sciences yet.

6.3. The elements of a dynamic microsimulation model

The basic principle of a time-driven microsimulation approach consists of formalizing change at the level of each individual, and therefore to consecutively review *all the individuals* at each considered step in the simulation. That given, the nature of the data involved and the formalization type of the individual change varies greatly from one application to another.

6.3.1. *The different sources of microdata: comprehensive information, samplings, artificial worlds*

When applying a microsimulation model, an important constraint concerns the initial situation, i.e. a group of individuals, each carrying a list of attributes. During the simulation, some individuals will disappear whereas others will be created and given a set of attributes within the model. In this way, even if the initial population corresponds to an observed situation, the number of entirely constructed individuals will increase throughout the progress of the simulation. The situations vary from one application to another and range from the use of entirely fictional data to the use of complete and comprehensive databases.

Certain applications rely on entirely artificial worlds. The goal is then to test the effects of different rules of individual behavior or of different initial configurations from a theoretical point of view, without attempting to reproduce a specific observed situation. Such is the case, for example, for the model developed by Basu and Pryor [BAS 97] in order to simulate the transition of a socialist economy to a free market economy. The initial situation is made up of 1,000 individuals and one government agency in which all agents receive the same salary. These individuals are different in their competitiveness and in their aspiration to become entrepreneurs. These attributes are determined by a random variable. Throughout the simulation's progress, certain individuals create companies, recruit employees and a diversified economy will gradually emerge. The model is entirely theoretical and uses fictional data in order to emphasize the effects of competition mechanisms in the economic activity structure.

Box 6.1. *Three models for one question, technical comparison*

In order to test the respective advantages of different forms of models, a short example was developed. The idea is to model the evolution of the Swedish population over 100 years, with the existing information on age, gender and municipality of residence of the individuals at the initial date of 1990. Let us suppose that we want to test different hypotheses on fertility behavior change. In Sweden, the observed current fertility index is 1.5 children per woman, although the national statistics bureau of Sweden (SCB) bases its predictions on the hypothesis that this rate will increase to 1.8 (as has happened). The smallest observed value until now for the longitudinal fertility (final descendants) is of two children per woman. If we only take into account the age and gender, we can use a traditional model of cohort relying on mortality and fertility tables by age. Considering spatial variations of demographic behavior will lead to use a model based on 2 genders, 100 ages, 280 municipalities, in other words corresponding to a three dimensional table of 56,000 cells. Three models were developed:

– a deterministic demographic model operating on the 56,000 cells: a mortality rate of 1% for 65 year-old men residing in municipality M, for example, signifies that out of 10,000 men who are 65 years old in this municipality, exactly 100 die each year;

– a stochastic demographic model operating on the 56,000 cells: the mortality rate follows a probability distribution associated with a Poisson error, and the same average rate of 1% then produces a random number of deaths, for example, 94 or 105 for 65 year-old men;

– a microsimulation model which reviews each of the 9 million individuals: each 65 year-old man from municipality M has a 1% probability of dying and a random draw is made to determine if that is the case.

With the SCB hypothesis, which predicts an increase of fertility of up to 1.8 in the next decade followed by a stabilization of this value, the total population will slightly increase until the middle of the 21st century and then will slightly decline to reach 8.8 million residents in 2090, a value close to 1990. On the other hand, if we presume that the former 1.5 observed value would remain the same in the future, the population of Sweden will decrease considerably during the next century. It will lose over three million residents and go down to 5.8 million residents in 2090. Such importance causes a difference of 0.3 in the fertility rate in the long run.

The three models converge to provide these results but the first one needs 4 mn of calculation time, the second one needs 8 mn and the microsimulation 1.5 h. Then, if there are no other attributes to consider, there is no reason to use the microsimulation to answer this question. On the other hand, if we want to consider hypotheses on fertility behavior differences based on education, revenue, home situation, then the number of attributes increases and we just need 6 or 7 in order for the number of cells to exceed the total number of individuals. In this case, the microsimulation becomes the better choice for performance in terms of calculation time and memory.

Other applications aim at basing the simulation on an initial situation with individual characteristics as close as possible to the observed population at a given time. For the team of geographers from Leeds (one of the few microsimulation teams that take into account the spatial dimension), the first step in the building of a microsimulation model consists of *producing* data at an individual level, from the knowledge of a series of cross-reference tables at a certain administrative level. In the first module of the microsimulation model, the individuals are reviewed one by one in order to be affected by coherent attribute values by using a Monte Carlo type procedure. Conditional probabilities according to the individual's other attributes are estimated from the cross-reference tables and the value of the new considered attribute[2] is then obtained by a random draw. This method will enable the building up of a coherent artificial population that has at an aggregate properties level similar to those for the actual population. The goal is twofold: on the one hand to produce at a fine scale an image of the spatial distributions of phenomena for which information does not exist at the desired scale (for example, on questions of revenue, health, taxation); and on the other hand to analyze the impact of a regulation change for individuals while considering the social and spatial contexts in which they live [BAL 01, BIR 95].

Spiekermann and Wegener [SPI 99] have, with a similar point of departure, attempted to take into account the effects of carpooling regarding home-to-work trips on energy savings in the Dortmund region. In these studies, the spatial dimension is also central. The authors use a GIS in order to determine the location of homes and jobs of the individuals from spatially aggregated information and a land-use map. The latter is used to establish a weighting system that is then formalized using a grid similar to the one in Table 5.2 (see Chapter 5). The residents and jobs are then located from a series of random draws. A micro database of the population of Dortmund (home and workplace) has thus been created with a spatial resolution of 50 m using aggregate information. The second step consists of also rebuilding the 210,000 home-to-work trips based only on aggregated level flows as information. Once these databases are built, the authors have developed an individual choice model concerning transport modes (individual car or different forms of carpooling). Several hypotheses were tested: rational attitude of individuals consisting of choosing a minimum detour or altruistic attitude consisting of favoring carpooling to maximize energy savings. The results of the model are given in numbers of car-kilometers in total and enable a comparison between the consequences of the different attitudes on energy consumption.

---

[2] Example: presuming that in county K, 30% of men in the age group between 40 and 50 are employed, 20% are blue collar workers, 10% are farmers, 10% are part of management, etc. When we need to determine the social category of individual male x, aged 43 and residing in county K, we will draw a number between 1 and 100. If the value drawn is between 1 and 30, we will give the attribute "employee"; if it is between 31 and 50, we will give the attribute "blue collar worker", and so on.

However, most microsimulation models use an initial situation corresponding to a sample of the concerned population. The data then comes from surveys which often provides a more specific description than the census data. They sometimes provide follow-up longitudinal information over a number of years. Several studies have been developed with the microsimulation tool to make forecasts in terms of family policy and retirement.

Finally, there are models developed from an initial situation reflecting observed comprehensive information on all individuals of the population. Such is the case with Fransson's [FRA 00] application concerning the evolution of the housing market in Gävle in Sweden. His objective was to show that in order to simulate the evolution of this sector, it is sufficient to only consider the sequence of individuals' actions. He highlights in particular the effects that the decisions of some will have on the others. The logic of the model is based on vacancy chains: a household that migrates provides a new potential residence that may satisfy the demand of another family. The model was tested on all 90,000 individuals in the municipality, with their family composition, their revenue, education and professional status as attributes and the 45,000 dwellings with their location, size, cost, occupation status and the attractiveness of the neighborhood as attributes. The individuals are located according to their family and each family is linked to a house. Globally the results of the model (occupation level of houses, profile of occupants, etc.) reveal a strong similarity with the observed evolution. However, an over-representation of children was registered, due to a different age structure of the immigrating population from that of the residing population. The model has emphasized the dissatisfaction of the demand from young couples and the obstacle that the shortages on the housing market represented for the start of their cohabitation.

The SVERIGE[3] model, developed by a research team in Kiruna and Umeå[4], includes the total Swedish population (approximately 9 million individuals). The access to a complete longitudinal database of the Swedish population from 1985 to 2003 has provided a solid empirical basis for this experiment [VEN 99]. We will come back to this model when we discuss applications (see section 6.4).

---

[3] SVERIGE: acronym for *System for Visualising Economic and Regional Influences in Governing the Environment*.

[4] At SMC: Spatial Modeling Center, research center at the Department of Social and Economic Geography at the University of Umeå, with a branch in Kiruna.

6.3.2. *Statistical procedures or agent type autonomy: the different ways to formalize individual change*

The simplest attribute of individual change in time-driven microsimulation is aging. For each step in time, each individual ages one unit of this time (months or years according to the applications). Age then is used as a driver in several rules, either in a deterministic way (mandatory age for school, retirement age, for example) or in a probabilistic way (fertility, mortality, for example). All microsimulation models have in common the fact that they simulate change at the basic level of the individuals, ,and most often use a Monte Carlo type procedure to determine individual outcome deviation from the expected value determined for that person in his/her current situation with the help of rules or estimated equations. This procedure consists of a random draw from a specific calculated distribution for each individual, at each time step, for each event that may happen socially or economically (birth, death, cohabitation, demographic mobility, loss or change of jobs, end of studies). The formalization of probabilities of transition from one state to another will vary greatly from one model to another.

The basic tradition is to model behavior using statistically estimated equations. The probability that a certain event will affect a certain type of individual in a certain situation is evaluated by discrete choice regression models, with parameters estimated using samples that are preferably the same or a larger than the start sample for the simulation. For example, the probability that a woman gives birth to a baby during a given time period will depend on her attributes in general, age, her matrimonial and professional status, her income, origin of birth, her education and sometimes, in models with memory, the year of birth of previous children. The main microsimulation models, CORSIM[5] in the USA and DYNACAN[6] in Canada, have been developed in this way. These models function well but their only spatial dimension corresponds to the state level [CAL 96]. Their main drawback is that they are made up of a very large number of equations so that the interconnections are sometimes difficult to control [WIN 00]. These models are very functional to provide specific impact assessment, of different political or economic changes. However, their ability to achieve plain accurate forecasts are not necessarily superior to models with a simpler specification. Their explanatory capability may be weak due to the high number of equations involved. Nevertheless, this is mainly a consequence of their complexity and ambition; replacing estimated equations with the same amount of *a priori* rules does not improve performance.

Other models focus on the understanding of the processes at the basis of the individual decisions. Rather than producing precise and reliable forecasts, their goal

---

[5] CORSIM: acronym for *Cornell Microsimulation Model*.

[6] DYNACAN: acronym for *Dynamic Microsimulation Model for Canada*.

is to show (or even discover) the consequences of different forms of interactions between the individuals. They are more often based on a qualitative logic, formalized with rules. Whereas in the previous case change is formalized based on estimated probabilities for annual events (birth, death, etc.) in a time-driven fashion (there is generally one and only one chance to have a child each year), here the models are often directly event-driven. Events or chains of events, trigger individuals to reach one or other decision depending on which environmental context they are in according to a set of causal hypotheses. This type of model was first applied in a simulation of 1,000 individuals over a 100-year period [HOL 89]. It produced individual biographies from the formalization of interactions derived from postulated assumptions about what each person *wants to*, *ought to*, *must* and *can* do in each choice situation. The idea was to model individual choice in a multivariate context with individualized behavior derived from a few basic theoretical properties of the person rather than estimated from observed behavior. For example, during his studies, an individual wants to follow a certain course of study. If this course does not exist in the city of residence, this wish will lead to a migration subjected to resource constraints, which are on the one hand individual (previous certifications and financial capacity) and on the other hand are social or collective (number of places available). This model was formalized from a series of logical rules and attempted to integrate certain concepts of *time-geography* from Hägerstrand (see Chapter 4 and [HAG 70]). The weak computer capabilities of the time in which it was created stopped its development.

Finally, among recent models called microsimulation models, some have been developed with the help of multi-agents systems (see Chapter 7). The emphasis in these models is on the communication capability between agents and the adaptation mechanisms. That is the case with the model discussed previously, simulating the appearance of a private sector in an initially socialist economy. The model includes three types of agents: the individuals, government owned companies and private enterprises. As with any microsimulation, the individuals are reviewed one by one. Depending on their character and a certain amount of chance, the individuals can change status: an employee in the public sector can start his own company (the communication system associated with MAS enables him to communicate with the other individuals in order to recruit employees) or be recruited in the private sector, an employee of the private sector can lose his job or change company. In addition, the individuals consume and choose the company that will supply them. Supply and demand mechanisms are thus introduced in the model, resulting in price adjustments although no optimization mechanisms have been explicitly introduced. The rules are defined at the individual level and their behaviors are diversified and liable to evolve in time following learning mechanisms. It is an experimental approach in which assumptions on the behavior of agents are made and the emerging structures are then observed at the macroscopicall level. Other authors have operated with the same logic, but by centering the core of the simulation on other types of processes, for example, on residential segregation in [POR 94], of cooperation between agents in hunter-gatherer societies in [KOH 00], etc.

In fact, the way in which individual behavior is modeled in microsimulation can refer to different theoretical frames, where the three extreme poles would be: 1) the main stream time-driven estimated micro-economic frame 2) an empirical frame only based on statistical regularities to determine the relation between the profile of the individual and the decision made; 3) the "agent" perspective where the decisions of the individuals are directly based on the nature of their interactions with others and where the decision making rules can vary through time according to these interactions, thanks to adaptation capabilities. Some models are firmly set in one or the other of these perspectives. However, many of them tend to be hybrid. Such is the case in most of the examples used in this section, and particularly those that will follow and which are based on the combination of statistical and rules logic. For instance, the rule of a probability equation with several changing individual and context drivers is precisely to adapt the individual response to changing conditions, but based on observations. Nothing prohibits certain events from changing the equation dramatically, creating a phase shift in the trajectory of the person and in the composition of social institutions emerging from a new set of relational properties of the model actors. The difference is more about purpose, exploring consequences of basic theoretical assumptions or making impact studies that at least faintly resemble observed distributions in an observed population. The connection between time-driven microsimulation and estimated equations is trivial; most longitudinal observables are produced on an annual base. It is harder to obtain direct estimates based on observables for durations and events occurring several times a year. Beyond differences in estimation ambitions and possibilities, the difference between an event-driven spatial microsimulation and an agent-based model is entirely rhetorical.

6.4. National forecasts and simulation of individual biographies with the SVERIGE model[7]

The SVERIGE model mentioned previously is a basically time-driven spatial microsimulation model integrating the entire population living in Sweden. The objective is to use a basic operational model to which we can add different modules centered on specific questions about society or environmental problems. In this way, the model will enable us to test scenarios with regard to the effects on the evolution of pollution produced by future behavior changes from individuals, for example, or

---

[7] Later development includes introduction of modules for different transfer payments as well as explicit labor supply and demand clearing.

to evaluate the impact of different political decisions (in terms of health or immigration, for example).

The core of the model is mainly demographic and its structure is similar to that of large North American models CORSIM and DYNACAN which have largely inspired it. The structure of the model is modular: fertility, mortality, cohabitation and marriage, divorce, leaving the family home, education, employment and earnings and immigration are the basic modules of these models. They needed to be adapted to a different cultural and institutional context from the North American context (the relations between men and women, policies on social equity, for example, are very different and imply the implementation of other rules) and to be calibrated to Swedish observations. In order to do this, the comprehensive data from the 1985 to 1995 period has been used. In addition, the spatial dimension, which is absent from North American models, has been explicitly introduced using a specific module focusing on the modeling of migrations as well as by making the individual decisions dependent on the spatial context in which they are. A module on emigration has also been added. These modules are sequenced in the microsimulation and change is formalized with equations, transition matrices and logical rules.

In this first example, the emphasis will be on different levels of production of results. Variants and extensions of the SVERIGE model are continually developed. A simplified version was used in this section in order to illustrate some kinds of outputs that this model can produce. This simplified model was applied to a sample of 14,000 individuals, who were randomly drawn from the file of the 9 million Swedish residents. The simulation lasted for 100 years, from 1990 to 2090. Around 20 attributes (age, education, activity, revenue, place of residence, etc.) are updated each year for each individual and the events associated with them (birth, death, start of education, change of employment, migration, etc.) are also stored.

6.4.1. *Classical aggregate outputs*

One form of output concerns the evolution of the population according to the different categories considered in the model. Figure 6.1, for example, simultaneously represents the evolution of the total population and its distribution based on the education level of the residents. We can globally observe a slight growth in the total population until the middle of the 21st century and then a decline. This global evolution is the result of hypotheses relative to demographic components (fertility, mortality, migration) used in the basic SVERIGE model scenario, corresponding to an increase in the cross-sectional fertility index to 1.8 and to a decrease in mortality which is higher during the first decades and then decreases over time. With regard to the distribution of the population by education level, the proportion of population reaching a college degree level increases considerably, whereas the proportion of less educated people rapidly decreases in the first decades to stabilize later.

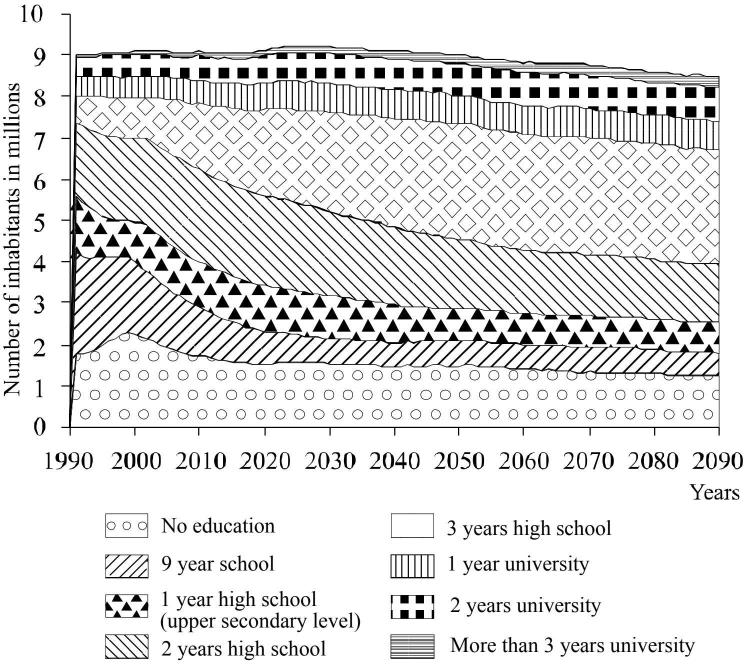


Figure 6.1. *Number of residents and education level: simulated evolution from 1990 to 2090*

The results can also be spatially aggregated. Figure 6.2 gives, for example, an idea of the redistribution of the population between the country's labor markets (108 regions delimited based on commuter flows). The rank-size representation of the labor markets in 1990 and in 2090 reveals a high stability in the hierarchical organization of the population. Beyond a multitude of individual migrations, the size relations between the employment centers remain relatively stable. This result converges with those of several dynamic models developed at meso-geographical level and corresponds to long-term trends that have been empirically observed for most settlement systems. The current distribution of the Swedish population at the parish level, for example, is explained at 50% by the population pattern in 1810 [HAK 00]. However, the slow speed of the simulated change by the year 2090 seems somewhat underestimated.

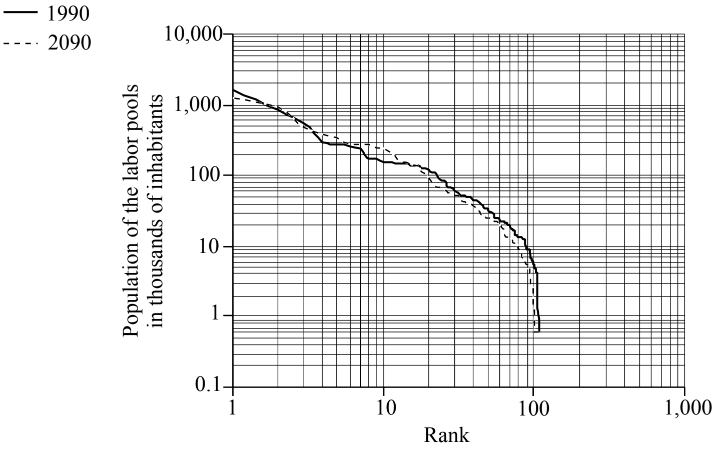
On the other hand, an examination at a finer scale reveals several changes in places relative to the poles within the hierarchy. At local level, the population variations are in fact more random and more sensitive to variations in the values of the model's parameters. The simulation results are certainly less reliable at this level, especially at the time scale considered. The prospective scenario is mainly based on the extension of individual migratory trends in the 1990 to 1995 period. The choice destination model of individuals that have chosen to migrate is gravitational. The distance from the place of origin, population mass and the activity rate of the destination are the main elements of the function[8] of choice. Long-term predictions would have been different if the parameters had been calibrated from behaviors at the end of the 1990s. We would then have obtained a more substantial increase of large cities and especially Stockholm. In fact, it does seem that the concentration of people in the larger cities of the national system increases during good economic times, whereas net migration is more favorable to regions that are away from the attraction areas of the large centers in slow economic periods. The effect of this economic logic is lessened by the fact that, according to a Swedish survey, only 20% of people migrate due to the labor market. Finally, we will note that natural growth has recently had a tendency to be more of a factor than migrations in explaining the growth differences between municipalities. These differences are linked to distributions of the population by age, which are themselves inherited from migrations of previous periods. The links between individual migration choices and society's regional and macroeconomic contexts are complex and cannot be formalized with a simple rule. The interest in such a microsimulation model is not to predict exactly the future distribution of the population, but to provide a helpful tool for our questioning, to test the long-term consequences of certain changes in the preferences or behaviors of individuals as well as to evaluate the impact of different forms of relations between the macroscopic, regional and individual levels.

6.4.2. *The biography of Kristina*

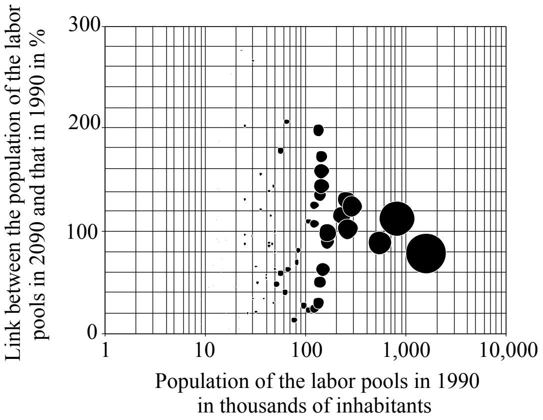
The aggregated form results presented above are traditional and generally correspond to the demand of decision makers. However, we can also focus the results on the specific individual trajectories and get closer to the level where the model is formalized. This does not really bring answers to experimental questions. However, such results can help us locate particular situations and specific relations that can trigger questioning on behaviors which are not very predictable *a priori* and that will mainly be used, on a practical level, to detect possible errors that would be difficult to identify from a simple review of aggregated tables and figures.

---

[8] Of multinomial logit type.



a) Rank-size distribution of Swedish labor pools in 1990 and in 2090



b) evolution of the population of the Swedish labor pool according to their initial size

Figure 6.2. *Simulated redistribution of the population between labor pools in Sweden between 1990 and 2090*

Table 6.1 shows the list of current events generated at individual level by the model and which lead to the transformation of certain of the individuals' attributes. In order to illustrate the sequence of events marking a biography, we propose to follow a female born in 1993 in Kiruna that we will call Kristina (see Table 6.2). Each time that an event concerning her occurs during the simulation, it is registered in the table with the corresponding date as well as Kristina's age at the time. This list gives us a basic biography of the individual through a set of identified elements. A first glance at the list gives the impression that the main task of the model is to locate income changes. In reality, the noted changes have very unequal values. In fact, for an employee, earnings change generally every year, sometimes modestly. Events such as birth, death, marriage or a migration are less frequent but have far reaching repercussions.

The list also shows that Kristina worked while studying. Such is sometimes the case but this fact may also reflect a mistake in the programming of the model. Certain errors are easier to identify by inspecting the individual biographies than by studying the results through aggregated tables. Our implicit knowledge of what is possible, appropriate, common, incredible or impossible for one individual, in a certain situation, is very wide. This non-formalized knowledge comes from a wealth of information on the individual in his context and reinforces validation of the model with respect to other traditional methods that refer to the aggregated level.

Kristina leaves Kiruna when she is 26 years old and experiences three more migrations between different labor pools. She goes through slightly more migrations than the current average in Sweden. In addition, she moves twice within the same labor pool, the first time in Kiruna, when she leaves her family home and the second time when she divorces. On the other hand, we will notice that her marriage does not coincide with a change in residence. Except for one year, she works throughout her adult life, simultaneously studying at certain times. She will have no children and die in Stockholm in 2077.

1. is born
2. immigrated
3. began an education (schooling, studies, training, etc.)
4. Finish his education
5. left the parental home
6. started a job
7. got married
8. had a baby
9. got divorced
10. her/his partner died
11. moved in the region
12. stopped working
13. died
14. changed his/her working hours
15. new income
16. emigrated
17. left the region

Table 6.1. *Types of events generated in the microsimulation for one person*

The locations are still tracked in a very crude way in the SVERIGE model. There are still no rules linking the decisions of individuals to the context in which they exist, with its opportunities and constraints (but is currently in development). A more detailed account of the environment of individuals would enable us to replace some simplified assumptions with estimated equations and logical rules, in other words substitute a concept driven model by a data driven model. This transition is done in a progressive manner; each step is submitted to a validation operation based on the coherence of individual trajectories as well as on the evolution of the population's macroscopic descriptors.

Year	Age	Events
1993	0	is born
2000	7	began her education
2009	16	started a job
		changed her working hours
		new income
2010	17	new income
2011	18	new income
2012	19	finished her education
		new income
2013	20	new income
2014	21	new income
2015	22	new income
2016	23	left the parental home
		moved in the region
		new income
2017	24	new income
2018	25	began her training
		got married
		new income
2019	26	new income
		left the region
2020	27	finished her training
		new income
2021	28	new income
2022	29	stopped working
		changed her working hours
		new income
2023	30	started working
		got divorced
		moved in the region
		changed her working hours
		new income
2025	32	new income
2026	33	new income
2027	34	new income
2028	35	new income
2029	36	new income
		left the region
2030	37	new income
		left the region
2031	38	new income
2032	39	new income
2033	40	new income
		left the region
2034	41	new income
2035	42	new income
2036	43	new income
2037	44	new income
2038	45	new income
2039	46	new income
2040	47	new income
2041	48	new income
2042	49	new income
2043	50	new income
2044	51	new income
2045	52	new income
2046	53	new income
2047	54	new income
2048	55	new income
2049	56	new income
2050	57	new income
2051	58	new income
2052	59	new income
2053	60	new income
2054	61	new income
2055	62	new income
2056	63	new income
2057	64	new income
2058	65	new income
		stopped working
		changed her working hours
		new income
2077	83	died

Table 6.2. *Simulated biography for Kristina*

6.5. A simulation of population spatial dynamics with MICDYN[9]

This second example[10] is based on the production of results and predictions at a finer spatial level: that of a French municipality[11]. The objective of the model is to test the impact of different hypotheses concerning job growth on the evolution of the population's spatial distribution at municipality level in the Hérault and Gard departments [ASC 00]. The area of study is made up of approximately 700 municipalities and a population of 1.4 million people in 1990, which is the date considered as the starting point in the simulation.

6.5.1. *Operation of the MICDYN model*

The idea behind the MICDYN model is to associate mechanisms operating at individual level with others that are clearly meso-geographical. We either introduce these mechanisms as constraints over the individuals' choices (perspective of *time-geography*; see Chapter 4), or we formalize certain mechanisms at meso-geographical level. This last alternative has been chosen and MICDYN is a microsimulation model associating micro-objects at two levels:

– the individuals: they are created from aggregated tables referencing different variable pairs according to a procedure similar to that described in section 6.3.1 [BAL 01]. The first step consisted of generating demographic attributes (gender, age and marital status) socio-economic attributes (social category, economic activity category) and location attributes (place of residence, place of work), and to assemble these individuals in coherent households;

– the municipalities: they are characterized by attributes derived by aggregates of individuals, neighborhoods (set of municipalities within *k* km, for example) and the existence or non-existence of certain services (school for example).

Figure 6.3 illustrates the way in which these two levels are associated in the model.

Most of the rules for change are formalized at the individual level. Each step in time corresponds to one year and all the individuals are considered one after the other in order to evaluate the changes affecting them during the year considered. These changes are determined in different ways:

---

[9] MICDYN: dynamic microsimulation acronym.

[10] The model explained in this section was developed within the ARCHAEOMEDES II project, directed by S. Van der Leeuw and financed by DG12.

[11] France is divided into around 36,000 municipalities.

– aging: the age attribute for each individual is simply increased by one unit each year;

– the occurrence of a certain number of events is determined by a Monte Carlo type draw: some are simple to maintain, such as mortality, the probability values are drawn from mortality table by age. For others, we do not have observed data tables. This is the case for the probability of leaving the parents, of meeting a partner, of separating, of starting higher education. Other events are still even more complex to formalize: the birth of a child depends on the "finding a partner" process and on the fertility by age tables;

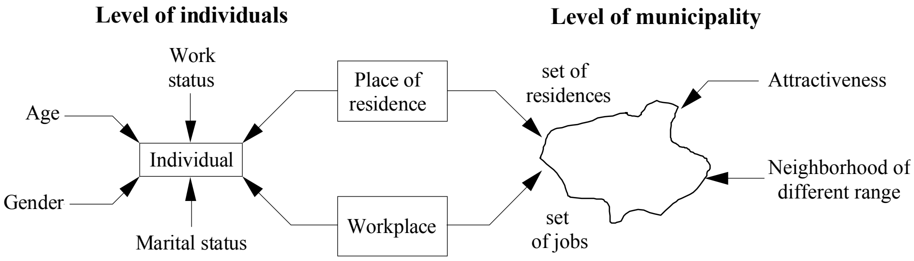


Figure 6.3. *Individuals, aggregates, spatial entities, neighborhoods (source: Aschan et al. [ASC 00])*

Possible mobility causes	Type of event	Necessity of finding a job	Necessity of finding a residence
Change in the family	Young leaving the parental home	Maybe	Yes
	Cohabitation	Maybe	No
	Separation	No	Yes
	Death of spouse	No	Maybe
Change in professional life	Immigrant	Maybe	Yes
	Loss of employment	Yes	Maybe
	End of schooling	Yes	Yes
No specific reason	Not integrated in the model	Yes	Yes

Source: Aschan *et al.* [ASC 00]

Table 6.3. *Events increasing the probability of a migration in the MICDYN model*

– the chain of events: the occurrence of certain events increases the probability of certain other events. The probability of migrating will be stronger for an individual finishing school and finding a first job, for an individual separating, for two individuals getting together as a couple, for an unemployed person finding a job, etc. (see Table 6.3). Even if the most frequent causes are family-based or professional, some migrations are for more specific individual reasons, which we have considered here by giving each individual, at each step in time, a probability (even if very low) to migrate, without any one element being the initiator;

– the arrival of new immigrants in the region: the object is to combine them in families, to give them a job corresponding to their profile and a residence.

6.5.2. *Determining workplaces and places of residence of migrants*

The most delicate part of modeling consists of deciding new places of residence and work for migrants, whether they come from outside or from the departments studied. In MICDYN's current version, the destination is determined by the availabilities of housing and employment corresponding to the social and occupational profile of the individual, as the residence and workplace are linked by a constraint of maximum distance. The empirical study shows that 90% of French drivers travel less than 40 km to reach their workplace [BER 98]. The 40 km threshold was also retained in the simulation to characterize the radius of the area surrounding the residence (respectively workplace) of the individual where the workplace (respectively residence) must be researched. In addition, during a workplace or residence selection procedure, the closer of the two large work centers of the region, Montpellier and Nîmes, is always included in the potential choices for the individuals, although with a low probability if these centers are farther than 40 km from the considered municipality. If travel distance is longer than 40 km, the individual will research a more appropriate residential location during the next iteration. All possible destinations are thus classified and the definitive choice results from a draw, which is based on the number of possibilities offered in the different areas. According to the events that have led to the migration, the base of the research is focused on housing and employment since the location of one is a constraint for the location of the other.

This basic mechanism is made more difficult when the individual studied is part of a couple, for example, in the case of a household where one of the partners is unemployed. For each iteration, an available job corresponding to his profile is researched in all the municipalities located at less than 40 km from the municipality of residence, as well as Nîmes or Montpellier. A random draw is used to determine the chosen employment. If the draw results in Nîmes (respectively Montpellier), the migration will occur only if the household finds available housing and if this new residence is not farther than 40 km of the partner's workplace. There is then a close interaction between the decisions of the individuals within the same household. In this version of the model, the migrations then become constrained by the availabilities of housing and employment. These availabilities are generated in an endogenous way, based on internal redistribution logic (certain individuals will quit their job, certain households will leave their residence) and in an exogenous way from an injection of new jobs and/or of new residences. An economic induction mechanism also appears in the model and is formalized at municipality level, as the increase of the population in a municipality leads to a certain increase in service employment.

The first simulations have systematically led to an overrepresentation of the population in the large centers and in the hinterland, and an under-representation in suburban areas. Simple adjustments on parameter values have not been sufficient to adjust the results. It has been necessary to clearly introduce a meso-geographical constraint in order to account for the diffusion of urban growth: when the density of the municipality population exceeds a certain threshold, the population increase is transferred to a contiguous municipality. Parallel to the rules defining change at individual level, this model contains then two rules defined at the municipality level, one accounting for induction mechanisms and the other managing a spatial diffusion process.

6.5.3. *Simulating the population evolutions 1990-2040*

Starting with observed distributions in 1990, the model was applied over a period of 50 years with a time step of one year. For the year 2040, we then have a population of individuals about which we know the gender, age, family status, education, sector of activity as well as residence and workplace. We can aggregate the information in different ways in order to bring various clarifications to change. The results are represented in the form of graphs showing the evolution of a given variable (total number of individuals or jobs of a specific type) according to time, or of maps giving an image of the future distribution of a certain category of population or of change registered in the distribution of a phenomenon. The residences and workplaces of all the individuals are listed for each date. We can then, for example, represent the evolution of distance of home-to-work commutes depending on the regions (see Figure 6.4) and map the portion of commuters at municipality level in 2040 (see Figure 6.5).

Once set up, the model can be used to test the different scenarios. The following example shows the very different consequences in 2020 of an exogenous increase in employment in the center of Montpellier (see Figure 6.6a) or in the center of Nîmes (see Figure 6.6b).

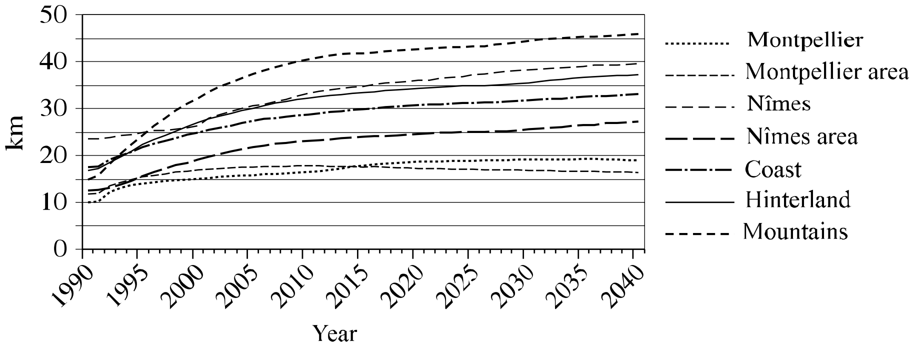


Figure 6.4. *Simulated evolution of the average distance traveled by commuters by sub-zone of Hérault and Gard over the 1990-2040 period*

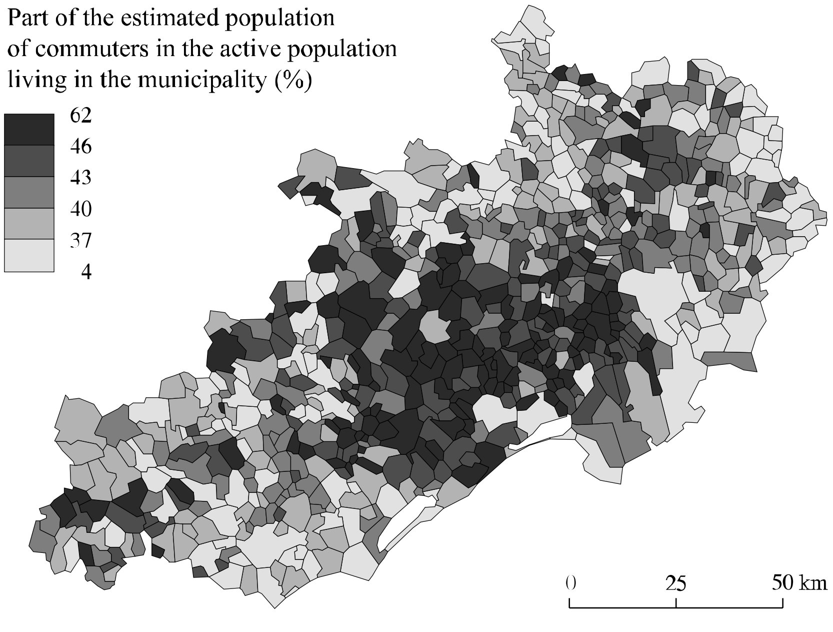
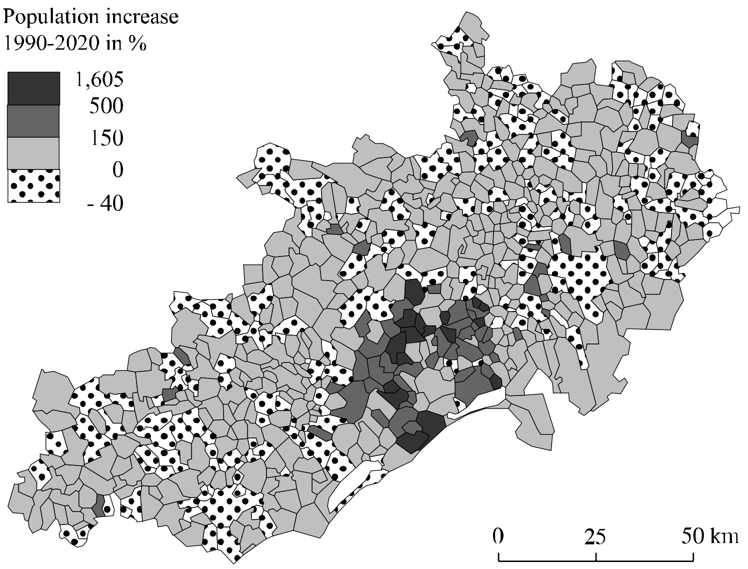
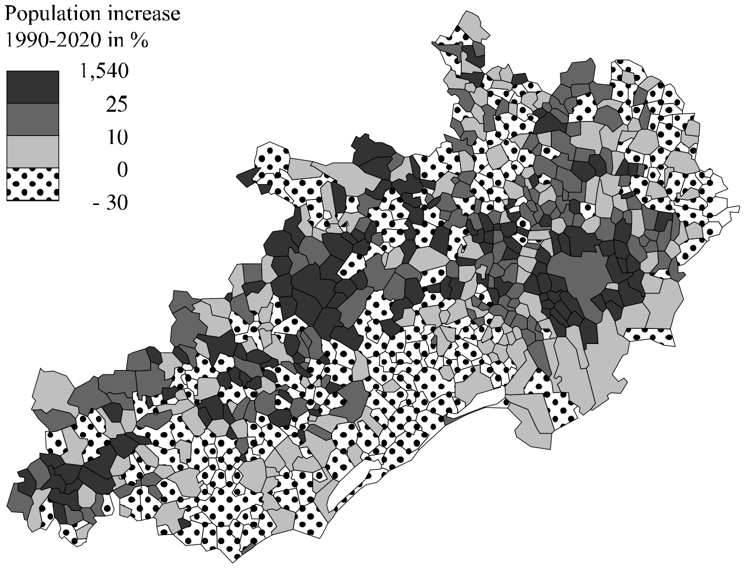


Figure 6.5. *Estimated proportion of commuters in Hérault and Gard in 2040*



a) Assumption of an increase in employment in the center of Montpellier



b) Assumption of an increase in employment in the center of Nîmes

Figure 6.6. *Consequences of an exogenous increase in employment on the evolution of the population between 1990 and 2020*

The investments in the center of Montpellier result in an important population boom in a very concentrated area of approximately 30 km around the city, to the detriment of Nîmes and the other municipality in the region. On the other hand, an exogenous injection of jobs in the center of Nîmes not only means an increase of its population, but also in several municipalities in the hinterland. Based on the local context, in terms of population density as well as in terms of the profile of resources, the effects of one exogenous action can be very diverse and do not always follow from intuition. The differences are due to questions of adequacy between the profile of jobseekers and the profile of existing activities. In fact, the exogenous employment profile created in Nîmes and Montpellier corresponds to the profile of the region as a whole, whereas the active population of Nîmes has a less diversified profile than that of Montpellier. In one case, these jobs correspond well to the profile of the active population and in the other case, less so. As a result, the effects of these job creations are of different nature.

6.5.4. *Perspectives*

The model presented above is actually based more on a prototype than on an operational tool. However, its development is already enabling us to reach several conclusions, especially from a methodological point of view. Our first point of discussion concerns the validation of the model. There has been no calibrating in the traditional sense of the word. Instead of defining difference criteria by term between observed reality and simulated evolution, the validation has relied on a qualitative evaluation of the plausibility of the evolutions at three complementary levels: individual, family and municipality. Given the frequency of the use of a stochastic element, each simulation will necessarily return different results. Logically, the model could not be used to predict exactly what will happen to the Dupont family or to the municipality of Saint-Dupargues during the next decades. Instead it will give us an idea of what can happen to families of the same type as the Dupont or to municipalities with the same profile as Saint-Dupargues. In addition, if the rules of change are defined for individuals, the results of the simulations must be interpreted at higher levels. On the one hand, we can question what will happen to categories of population: evolution of the number of farmers in wine producing towns, evolution of the number of single-parent families in the suburb of Montpellier, proportion of retirees in the coastal zone; evolution of the age structure of managers; increase in the length of home-to-work commutes, for example. On the other hand, we can question spatial regularities and the dynamics of certain town categories: evolution of small towns in the hinterland, evolution of employment in central business districts, expansion of suburban zones, intensity of future gradients in the spatial organization of growth, change in the space of certain discontinuities, evolution of intermediate zones, for example.

Simulations will help evaluate the differences in the form and rhythm of the diffusion of urban growth in the neighborhoods of Montpellier and Nîmes. They will also enable us to understand which factors can reinforce the decline of the hinterland or, on the contrary, lead to a demographic revitalization. On the other hand, they do not always help us interpret future differences between two similar and neighboring towns.

We still need to study the sensitivity of the parameters before using the model to test more sophisticated scenarios on the effects of new transport infrastructures or changes in the preferences and behaviors of individuals in terms of residential choice.

6.6. Conclusion

The applications presented in this chapter give us a good overview of the advantages and disadvantages of microsimulation modeling. On the negative side, we will remember a certain complexity in the design and calibrating problems, but these are part of all dynamic models used in social sciences. On the positive side, we can emphasize the freedom to produce results and therefore elements of questioning on a large variety of geographical levels once the model has been completed. The result is huge possibilities of use in various sectors, but always on the condition that the core of the model is adjusted correctly. We must emphasize that today the choice of this type of formalization does not clash with technical constraints due to the level of technological development. It is much more a philosophical choice [SAN 99, WIN 00] as is indicated from the discussion in the first sections.

In order to advance theoretically, several directions are open. To avoid a restrictive choice between two schools of modeling, one operating at meso-geographical level and the other with individuals, we probably should combine in a more systematic way the two approaches and think about their pairing. As we have demonstrated with the example of the MICDYN model, it is possible to integrate rules functioning at an aggregated level in a microsimulation model. In this example, the two operation modes are juxtaposed and it would be beneficial to integrate them even more. In this way, the aptitude of one municipality to attract new migrants depends on its *attractiveness*, term that we normally use for an aggregated level, from its relative position, its population, its equipment level and from its other various services. Each of these factors has a different weight, according to the individual considering them. A family with young children does not perceive the attractiveness of a place to live in the same way as an elderly person living alone. There would therefore be a lot of work to do on the measure of attractiveness, meso-geographical notion, modulated to the household categories and to individual decision logic. Certainly, such an approach would give a better indication of the interaction between the decision and the context in which the decision is made.

6.7. Bibliography

[ALL 97] ALLEN P., *Cities and regions as self-organizing systems; models of complexity*, Gordon and Breach Science Publishers, Amsterdam, 1997.

[ASC 00] ASCHAN C., MATHIAN H., SANDERS L., MÄKILÄ K., "A spatial microsimulation of population dynamics in Southern France: a model integrating individual decisions and spatial constraints", in *Applications of Simulation to Social Sciences*, G. Ballot, G. Weisbuch (eds.), Hermès, 2000.

[AXT 02] AXTELL, R. L., EPSTEIN, J. M., DEAN, J. S., GUMERMAN, G.J., SWEDLUND, A. C., AND HARBURGER, J., CHAKRAVARTY, S., HAMMOND, R., PARKER, J., PARKER, M., "Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley", in *Adaptive Agents, Intelligence, and Emergent Human Organization: Capturing Complexity through Agent-based Modelling,* B. J. L. Berry, L. D. Kiel and E. Elliott (eds.), *Proceedings of the National Academy of Sciences of the USA,* vol. 99 (Suppl. 3), p. 7275-7279, 2002.

[BAL 01] BALLAS D., CLARKE G., "Modelling the local impacts of national social policies: A Microsimulation Approach", *Environment and Planning C, Government and Policy*, 2001.

[BAL 00] BALLOT G., WEISBUCH G. (ed.), *Applications of Simulation to Social Sciences*, Hermès, Paris, 2000.

[BAS 97] BASU N., PRYOR R., Growing a Market Economy, Sandia Report, Albuquerque, NM 87185-0619, 1997.

[BER 98] BERROIR S., Concentration et polarisation: vers une nouvelle organisation des espaces urbains, Thesis, Paris University 1, 1998.

[BIR 95] BIRKIN M., CLARKE G., "The generation of individual and household incomes at the small area level", *Regional Studies*, 23, p. 553-548, 1995.

[BOM 04] BOMAN M., HOLM E., "Multi-agent systems, time geography, and microsimulations", in *Systems Approaches and their Application, Examples from Sweden*, Kluwer, 2004.

[BUR 96] BURA S., GUERIN-PACE F., MATHIAN H., PUMAIN D., SANDERS L., "Multi-agents system and the dynamics of a settlement system", *Geographical Analysis*, vol. 28, no. 2, p. 161-178, 1996.

[CAL 96] CALDWELL S.B., KEISTER L.A., "Wealth in America: family stock ownership and accumulation, 1960-1995", in *Microsimulation for Urban and Regional Policy Analysis*, Clarke G. (ed.), Pion, London, 1996.

[CAM 86] CAMAGNI R., DIAPPI L., LEONARDI G., “Urban growth and decline in a hierarchical system: a supply-oriented dynamic approach”, *Regional Science and Urban Economics*, vol. 16, p. 145-160, 1986.

[CLA 87] CLARKE M., HOLM E., “Microsimulation method in spatial analysis and planning”, *Geografiska Annaler*, vol. 69B, no. 2, p. 145-164, 1987.

[CLA 97] CLARKE K., HOPPEN S., GAYDOS L., “A self-modifying cellular automaton model of historical urbanization in the San Fransisco Bay area”, *Environments and Planning B: Planning and Design*, vol. 24, no. 2, p. 247-261, 1997.

[CLA 96] CLARKE G.P. (ed.), *Microsimulation for Urban and Regional Policy Analysis*, Pion, London, 1996.

[FER 95] FERBER J., *Les systèmes multi-agents: Vers une intelligence collective*, InterEditions, Paris, 1995.

[FRA 00] FRANSSON U., “Interrelationship between household and housing market: a microsimulation model of household formation among the young”, *Cybergeo*, no. 135, 2000.

[GIL 99] GILBERT N., TROITZSCH K.G., *Simulation for the Social Scientist*, Open University Press, Philadelphia, 1999.

[GUP 07] GUPTA, A., HARDING, A., “Introduction and overview”, in *Modeling our Future – Population Ageing, Health and Aged Care, International Symposia in Economic Theory and Econometrics,* Volume 16, Elsevier, 2007.

[HOL 04] HOLM E., HOLME K., MÄKILÄ K., MATTSON-KAUPPI M., MÖRTVIK G., 2004, *The Microsimulation Model SVERIGE; Content, Validation and Applications*, SMC, Kiruna, Sweden (www.sms.kiruna.se).

[HOL 07] HOLM E., LINDGREN U., HÄGGSTRÖM LUNDEVALLER E., STRÖMGREN M., SVERIGE, in Gupta, A., Harding, A (eds.) *Modeling our Future – Population Ageing, Health and Aged Care, International Symposia in Economic Theory and Econometrics,* Volume 16, Elsevier, 2007.

[HÄG 70] HÄGERSTAND T., “What about people in regional science?”, *Papers of the Regional Science Association*, 24, p. 7-21, 1970.

[HÅK 00] HÅKANSSON J., Changing Population Distribution in Sweden – Long Term Trends and Contemporary Tendencies, GERUM, no. 1, 2000.

[HOL 89] HOLM E., MÄKILÄ K., ÖBERG S., Tidsgeografisk handlingsteori; att bilda betingade biografier (A theory on behavior based on “time-geography”; building coherent individual biographies), GERUM, Report 8, Umeå, 1989.

[HOL 00] HOLM E., LINDGREN U., MALMBERG G., “Dynamic microsimulation”, in *Spatial Models and GIS: New Potential and New Models*, *GISDATA 7*, Fotheringham S., Wegener M. (ed.), Taylor & Francis, p. 143-165, 2000.

[KOH 00] KOHLER T., GUMERMAN G. (ed.), *Dynamics in Human and Primate Societies; Agent-Based Modelling of Social and Spatial Processes*, Oxford University Press, 2000.

[OSU 00] O'SULLIVAN, D., HAKLAY, M. "Agent-based models and individualism – is the world agent-based?" *Environment and Planning A*, 32 1409-1426, 2000.

[PAV 94] PAVE A., *Modélisation en biologie et en écologie*, Aléas, Paris, 1994.

[POR 94] PORTUGALI J., BENENSON I., OMER I., "Socio-spatial residential dynamics: stability and instability within a self-organizing city", *Geographical Analysis*, 26, p. 321-340, 1994.

[PUM 89] PUMAIN D., SANDERS L., SAINT-JULIEN T., *Villes et auto-organisation*, Economica, Paris, 1989.

[ROY 96] ROY G., SNICKARS F., "CityLife: A Study of Cellular Automata in Urban Dynamics", in *Spatial Analytical Perspectives on GIS*, *GISDATA 4*, Fischer M., Scholten H., Unwin D. (eds.), Taylor & Francis, p. 213-228, 1996.

[SAN 97] SANDERS L., PUMAIN D., MATHIAN H., GUERIN-PACE F., BURA S., "SIMPOP: a multi-agent system for the study of urbanism", *Environment and Planning B*, vol. 24, p. 287-305, 1997.

[SAN 98] SANDERS L., "Les décisions individuelles dans les modèles dynamiques", *Géopoint*, special no. "Décision et analyse spatiale", p. 294-299, 1998.

[SAN 99] SANDERS L., "Modelling within a self-organising or a microsimulation framework: opposite or complementary approaches?", *Cybergeo* (http://www.cybergeo.presse.fr), 1999.

[SPI 00] SPIEKERMANN K., WEGENER M., "Freedom from the Tyranny of Zones: Toward New GIS-based Spatial Models", in *Spatial Models and GIS: New Potential and New Models*, *GISDATA 7*, Fotheringham S., Wegener M. (eds.), Taylor & Francis, p. 45-61, 2000.

[WHI 97] WHITE R., ENGELEN G., "Cellular automata as the basis of integrated dynamic regional modelling", *Environments and Planning B: Planning and Design*, vol. 24, no. 2, p. 235-246, 1997.

[WIN 00] WINDER N., "Modelling within a thermodynamic framework: a footnote to Sanders (1999)", *Cybergeo* (http://www.cybergeo.presse.fr), 2000.

Chapter 7

Multi-agent Simulations of Spatial Dynamics

7.1. Introduction

Implementing multi-agent simulations means designing, building and modeling a complex system made up of individualized entities having a certain degree of autonomy and interacting with each other. Such an approach is proposed today for the resolution of artificial intelligence problems as well as for representing economic, ecological, geographical or physical processes. It is used increasingly in the explanation of spatial dynamics.

The studies based on multi-agent simulations are numerous. We can cite some pioneer applications from the 1990s such as Cinefil in parasitology [PIC 91], SimDelta in halieutics [BOUS 93], Manta in animal ethology [DRO 95], Sealab [LEP 96] in marine ecology and Simpop [BUR 96] in urban geography. All these applications have contributed to establish connections between the disciplines involved and computer science. The examples below will illustrate the diversity of modeled fields. All of them give an important place to the space in which the studied dynamics are deployed.

In the field of physical dynamics, Rivage [SER 00] is a simulator of water run-off processes. Research from Teles [TEL 98] focuses on multi-agent simulation of the alluvial plain genesis. Research from Breton [BRE 00] proposes a multi-agent method to calculate the network of forces in a pile of sand at equilibrium, with a problem resolution perspective (ecoresolution).

Chapter written by Jean-Pierre TREUIL, Christian MULLON, Edith PERRIER and Marie PIRON.

In the marine ecosystems and halieutics fields, Ichtyus [MES 94] was programmed for the study of gregarious behaviors and school formation by asking the question: what minimal hypotheses on the kinetic behavior of fish can explain the observed behavior of schools? As for Osmose [SHI 98], it is a simulator dedicated to the study of the properties of a marine ecosystem in the light of hypotheses on the modeled behaviors of predation and migration at the school level. Finally, Marlon [MIL 99] is a simulator dedicated to the study of the efficiency of different fishing techniques according to the spatial aggregation of the resource.

In the terrestrial ecosystems field and their management under pressure of various economic activities, we were able to simulate the effect of the different strategies of pasture from sheep farmers over the overgrown state of the rural landscape [LAR 98]. The studies of Bousquet *et al.* [BOUS 01] on the simulation of the hunting territory of a forest village of Cameroon analyze the way in which the current hunting activity organization constitutes, in fact, one way to manage the ecosystem. Mathevey [MAT 00] studies the interactions between the environment, the users and the territories in the context of the wet areas of Camargue. Other studies involve a model for an integrated management of the interior delta of the Niger river in Mali [KUP 99].

In urban research, we can mention the MICDYN [ASC 00] model whose objective is to understand and to predict, through a model of individual and household behaviors, the evolution of the weight of the different cities within a region. Portugali, followed by Benenson [BEN 98] designed intra-urban dynamics simulators to explain the interaction of changes of residence with the changes of status and social behaviors of households and the emergence of new forms of urban cultures.

Many simulators based on the multi-agents approach have been developed in many domains. Let us quote in biology, BioFilm [KRE 01, PIC 04, YER 06], a model of operation of bio-reactors, in ethology, models developed around the emergence of the ratios of predominance in the companies of primates (DomWorld) [HEM 04, 05]; in economics, models evaluating the impact of strategies of concurrent makes on the choices of the consumers like Cubes [LAM 02]; in urban research, the EuroSim simulator [SAN 06] is intended for the study of the evolution of the European urban system in continuity with SimPop; the Miro project [BAN 05] models intra-urban daily mobilities and ArchiSim studies the questions related to the road traffic and the behavior of the drivers [MES 06, DON 06].

In this chapter, we will first present the origin and the many dimensions of the multi-agent approach as it has been developed in computer research (see section 7.2). Then, by developing two different questioning examples, one based on the ecosystem model of the interior delta of the river Niger, and the other on the mathematical models of density-dependent phenomenon and their relation with the studies on runoff, as implemented in the Rivage simulator, we will analyze the different concepts involved in the modeling of spatial dynamics (see section 7.3). Finally, we will conclude with an attempt to propose a global point of view on the specific contribution of the multi-agent approach that these models claim to follow (see section 7.4).

7.2. The multi-agent approach

In 1989, Durfee [DUR 89] defined a multi-agent system (MAS) as a network loosely coupled with entities acting together to resolve problems that were beyond their individual capabilities. This definition linking MASs with distributed artificial intelligence (DAI) marks a decisive moment in a long and complex history [FER 95].

The MASs actually concern several research fields: the DAI obviously[1], but also artificial life, robotics, programming languages and distributed systems. It is difficult to qualify research on MASs as purely technological as it is too influenced by the interaction of numerous disciplines, from organizational and economic sciences to ethology and even physics. The history of MASs remains embedded in the circulation of metaphors which are often emphasized [LEN 94, BAT 96] within several scientific communities:

– a community of computer researchers studying new design methods for distributed artificial intelligence, man-machine network communication and distributed robotics where the management of interactions between agents, multiple computer processes and the program structures are very often inspired by metaphors of human or biological society organizations;

– a community of researchers from different disciplines, analyzing the dynamics of complex systems whether natural or social from simulations involving multiple interacting agents which, in these simulations, represent identified entities in the real world.

There are two different arguments on agents and multi-agent systems which are closely interrelated: the first is a technological argument concerning a way to view the interaction between processes that execute programs, and it results on research that is relative to *multi-agent systems*. The second argument aims for a general understanding of the interactions where, to borrow the definition from Shoam, an agent "is an entity operating in a continuous and autonomous way in an environment in which other processes occur and in which other agents exist"

---

[1] With which some countries have a tendency to become confused, particularly in the USA.

[SHO 93]. This conception is used in research where multi-agent systems are used as a tool of representation of the reality through what we call *multi-agent simulations*[2].

7.2.1. ***Multi-agent systems***

In this section we focus on the technical and computer science viewpoint of notions of agent and multi-agent system. This enables us to avoid definitions that are too vague and lead to confusion among different scientific fields [WOO 95].

Computer processes and agents

The MASs parallel a computer process, i.e. the execution of a program, and the life of a social being that must satisfy some constraints in his communications with other similar social beings and with an environment that he modifies and which in return, interacts with him. They completely develop this analogy by formalizing it and by attempting to apply it to the design and execution of software applications in different areas: problem resolution, industrial process or network operations control, knowledge base use, etc. Over electronic substrate material, they bring artificial worlds to life where autonomous agent populations perceive, think about, communicate, act and interact within an environment, all the while respecting constraints such as the pursuit of a collective or individual goal and maintaining relations demonstrating social organization between each other.

Typology of agents

In a common classification, we observe *reactive agents* (with a simple individual behavior, and whose interactions with other agents are also simple; these are agents that we can use in large numbers), *cognitive agents* (whose behavior is explained by elaborate representations of themselves, of other agents and of their environment, which leads to a utilization capacity reduced to a small number of agents). This classification is mainly a language convenience, established on an axis that will go from reactive agents to cognitive agents and which makes it possible to place the reality of the implementations. Other typologies have been proposed, trophic agents *versus* hysteretic agents [FER 95], hedonic agents *versus* eductive agents [BOU 95], to structure what is in reality a multidimensional space: dimension of the agent's memory (what can it memorize?), dimension of the perception (what does the agent perceive?), dimension of the internal control (what degree of freedom does the agent have as regards the choice of its actions or the modification of its basic behavior?), dimension of the communication (can the agent communicate? How? What types of

---

[2] In this chapter, MAS is reserved for multi-agent systems in order to differentiate them from multi-agent simulations that are a part of MAS.

messages can it address?), dimension of the organization (who can the agent interact with?), etc.

The work devoted to MASs supports concepts of autonomy, cooperation and organization and it attempts to clarify their meaning: the agents are *autonomous* entities, they interact in a way that is often qualified as *cooperative*, however, these interactions are limited by the *organization* of the system; this organization is in constant control of who can do what, who can interact with whom, and this limits the degree of freedom of the agents. We will focus on the first two concepts.

The concept of autonomy

One explanation of autonomy is the individualization of entities, the institution of a break between the agent and its environment. It cuts through the notion of *locality* already taken into account in the object oriented program. We find one specific formulation in the work of Kiss: "An agent is a local process which includes a private domain, on which it is the only operator, and it does not share with any other process" [KIS 92]. This break with the environment also manifests itself on a temporal level: an agent is a process that interacts only sporadically with its environment and who only functions, for the rest of the time, as an internal process, within its private domain and as a more finite temporal resolution. An agent is therefore capable of silent actions, opaque to the observer, that are not instant reactions to the environment's changes, nor actions that will instantly modify this environment. This autonomy can be more or less stressed, as the degree of autonomy is linked to the importance of the private domain, to the complexity of internal processes [KIS 92], to the existence of a rhythm of time and to a desynchronized operation in relation to the environment.

In a second explanation, to really be autonomous means being able to face alone and successfully the problems that the world puts in our way. This autonomy is linked to the viability of an agent, to its capability of adaptation, of survival in an environment that is ever changing and unpredictable [BOU 95, KIS 92]. The degree of autonomy of an agent is measured by its degree of success, its performances in a game with/against its environment, where the stakes (survival, respect of some constraints, reaching a specific objective) as well as the problems (unpredictability of the environmental dynamics) can be quantified.

These characteristics of autonomy are linked because the first one (the agent/environment decoupling) is the condition of the second (capacity of adaptation).

The concept of cooperation

The concept of cooperation is also the subject of several approaches [BRA 95]. According to Ferber [FER 96] and Drogoul [DRO 97], cooperation can be analyzed in two ways:

– we either concentrate on the agent and focus on the definition and formalization of a cooperative behavior. A cooperative behavior is then associated with some internal agent structures interpreted in terms of common objectives, commitments, agreements [JEN 93], which restrain the actions that they can individually take;

– or we use the observer's viewpoint, considering the system and its operation from the outside. A cooperative operation is then associated *a priori* to a reference by the observer of a condition to be fulfilled by the system; a condition that must be satisfied by the sequence of interactions between the agents. This condition can be the resolution of a problem that is assigned to the system as a task, the maintenance of some properties in time, or the permanence of the system in an uncertain environment.

The two viewpoints are not incompatible and complement each other as causal explanation and functional explanation [DRO 97]: the first one focuses on the analysis of the mechanisms that determine the agent behavior and actions within the agent itself; the second focuses on the characterization of the agent's own functional behavior independently of the mechanisms that trigger it. In reality, they have their roots in the distinction previously mentioned between cognitive, rational or intentional agents, which come from psychological and social metaphors (first point of view), and reactive agents which come from biological and ethological metaphors (second point of view).

Research on architectures and formalisms

There has been a lot of emphasis on the links between the concept of "object", as it is used in object oriented coding, and the concept of "agent"[3]. We must consider – from a computing point of view – the notion of object as an encompassing notion: an agent is an object that verifies some given conditions. As with any object, it is made up of attributes and methods and it is stuck in an inherited hierarchy. It is really an agent if its attributes, its methods and the way in which everything works respond to some specifications[4]. It is important to identify the general nature of these specifications that "make the difference".

---

[3] There is computer research on objects and other "concurrent objects", "players", etc.

[4] For example, an "asynchronous" message taken into consideration, a structuring of the group of attributes and methods in modules with a generic function (mailbox, memory, modules of perception, action, communication, etc.).

Studies [AKN 98, BRO 86, OCC 97, WER 96] concerning *the internal structure and operation of reactive agents* focus on the disintegration of the agent into a specific number of modules with particular functions – perception modules, effectors, deliberation modules – and the specification of the relations between these modules.

Research on *the internal structure and operation of cognitive agents* leads to formalisms[5] that enable us to specify the understanding of an agent, its beliefs, its mental states (goals, intentions, commitments), its competencies, its way of inferring from this or that specification choices concerning communication and action, of reasoning in uncertain situations, of building and presenting plans.

The *interactions between reactive agents* are often shown in an indirect way, through modifications made to the environment. The observation of social insect behavior has contributed in this field to a few simple but particularly productive ideas concerning the implementation of implicit coordination between agents. Computer tracking systems (the equivalent of pheromones) have been used to solve optimization problems. These interactions between reactive agents, through the environment, have been the subject of mathematical formalizations [SCH 97]; the track is here a spatial information field generated by the agents being and having its own diffusion dynamics.

The interactions between cognitive agents are often represented by formalisms that come from linguistics (pragmatics, theory of language acts [BOUR 92]). The idea is to describe conversations between agents in a social network as well as their goals. The agents must have a language enabling them to identify each other, to indicate their competencies, to require or provide information or services, to propose goals, to negotiate and to make commitments, etc. A particular effort has been made to identify the structures of typical conversations – interaction protocols such as *Contract Net Protocol* (CNP) which can be used in several different applied contexts.

The issue of the *organization* concerns the distribution of the agents' rights: who can do what?; who has access to what?; who can communicate with whom and about what? The importance of this notion is emphasized by [BAE 96], but there are few works devoted to it in MAS literature. Ferber [FER 98] describes the effect of the basic direction of MASs: the agents are autonomous, as little pressured as possible by social structures *a priori*, such structures will emerge in fact *a posteriori* from competencies of the agents and from their interactions. Two notions, however, have appeared early and have been reused and formalized: the notion of role

---

[5] These formalisms are based on extensions of the logic: modal logic, temporal logic, logic of possible modes.

[COL 96, FER 95] and the notion of group [MAR 90]. The Aladin system and its Madkit implementation [GUT 98] bring together the concepts of agents, of group and of role in the same interactions control architecture. In such systems, an agent can belong to different groups; its possible actions as part of the group are set by the role that it plays in this group. Efforts have recently been made with SMA where several layers of organization can coexist, following the example of what is happening in human societies [SER 00]. In these MASs, a group of agents can act as one agent with its own behavior rules and assert a certain control on the behavior of its members. The system must then incorporate the rules of the dynamic relationship between these different layers of organization (constitution and dissolution of the groups, prevalence of the rules of the group over those of the first level agents, for example).

7.2.2. *Multi-agent simulation of natural and social phenomena*

Early in the 1990s, the notions of agents and of multi-agent systems have interested researchers from natural and social sciences because they enabled them to make virtual experimentations and simulations without the obligatory mediation of mathematics. The source of this interest was in the large representation potential of these systems and in the possibility of using the potential for programming artificial worlds where a multitude of various entities interact: possibility of differentiating the action of agents from the reaction of their environment and the possibility of giving agents complex behaviors in the form of rules combining quantitative and qualitative aspects. It is in this context that the examples in the introduction are placed.

Modeling of interacting entities: an update of an old scientific trend

The investment of MASs in the simulation world joins a much older scientific trend which finds here a new way of expression and updating. This trend consists of an approach of modeling the world which directly takes into account the basic entities of a specific organization level in order to describe the interactions and to reconstruct from them the dynamics of higher organization levels. Starting from specific disciplines like physics and economics, it tends to cover a large scientific field with varied terminologies. We will talk about particular approaches in physical sciences [HOC 98], of individual-based modeling in ecology [GRI 99], of microsimulation in demography and geography [VAN 97] or methodological individualism in sociology [DUP 92].

The popularity of this scientific trend attentive to the relations between organization levels is partly linked to the simulation possibilities of information technology. However, it is also linked to the existence of mathematical tools – such as the theory of renormalization[6] [LES 95, WIL 92] – and of a group of theoretical questionings on the relations between organization levels which have as common origin statistical physics. In certain disciplines like ecology or demography, this trend is the subject of evaluation regarding its practical and theoretical contribution [GRI 99]:

– practical contribution first, when the particle-based or individual-based approach is the most immediate way[7] to model a phenomenon with the representation of the behavior of concrete elements (particles, individuals) at the organization level where this representation is the easiest;

– theoretical contribution then, when the determining intervention (over the dynamics of the phenomenon) of heterogenity of individual behaviors and situations is emphasized, particularly the differentiations caused by the spatial locations.

Using multi-agent systems in simulation: the "agentification" problem

The success of the use of MASs in simulation is initially based on the adequacy of their structures with a conceptualization of reality in the form of interacting entities and processes. However, such an adequacy characterizes in a more general way "object oriented programs"[8]. As long as the research on MASs can define the specificity of agents in relation to the objects, the question then becomes knowing what we can "agentify" and why? We can, in the first place, show human actors as agents (individuals or groups, institutions) or even human beings and keep the term object to represent the entities of the environment. We then use[9] fully – and in the opposite sense so to speak – the metaphors on which the specific agent structures and the systems they form have been built.

We can make a different choice based on considering the computer architecture of the simulator and not the nature of the entities to represent. In this spirit, we show an entity as an agent (an individual, a group, or even an agricultural piece of land, a region, an element that is fluid, a river), as long as it seems useful to do it *electronically*. The idea then is to code, in the form of agent structures, the perception of the environment, the understanding of some mechanisms and, in a

---

[6] As a calculation technique for upscaling in critical situations where an event at the smallest scale will tend to affect all the higher scales.

[7] By which we mean not spread by more abstract notions. In this way the immediate representation of a population is that of a group of individuals distributed in a space, and not that of a distribution of density.

[8] The first object oriented language, SIMULA, was designed as its name indicates for simulation purposes.

[9] Or perhaps "we could use", since the practice of programming simulators tends to distance itself from this ideal, except when using platforms which pressure the user into respecting their specifications.

broader way, the *tools necessary for ensuring the control of the correct progression of the simulation*. In the most extreme case, the agents of the simulator may not correspond to any of the entities of the simulated world and only be devices representing the scientists themselves and their knowledge on the implemented processes and their constraints. We are then in the "metaphor of the puppeteer" implemented in some simulators [BOUS 93, FIA 98].

7.3. Modeling of spatial dynamics

The consideration of space in the analysis and the modeling of phenomena that take place in it are not straightforward. The history of disciplines as diverse as economy, sociology, ecology and hydrology shows two main attitudes:

– the first one tends to describe mechanisms or processes, to characterize dynamics independently from space. The integration of space in the theories and models is reserved for a refinement step. We can qualify this attitude as a functional approach;

– the second one on the other hand tends to favor the description of the anchorage of the phenomena in space. By nature, it is the objective of geography to bring out and explain the spatial structures manifesting this anchorage. However, what we qualify here as spatial approach is found in other disciplines.

The concern for considering space in disciplines where it is not *a priori* their main goal has been made evident when the non-spatial theories and models failed to take into account observed dynamics, even when these dynamics were comprehended with non-spatial indicators. On the contrary, the disciplines of geographical orientation have been confronted with the necessity to describe, characterize and explain the evolutions of spatial structures. Two movements have emerged:

– the first one consists of spatializing functional approaches;

– the other consists of dynamizing spatial approaches.

We should then speak of spatial dynamics and dynamic spaces. We will group both expressions under one term: "spatial dynamics". We will start by analyzing two modeling approaches for these dynamics, referencing both of them to interactions between agents. The first one is more "computerized" (see section 7.3.1) as it is based on the construction of a virtual world and its exploration by simulations and the second is more "mathematical" (see section 7.3.2), as it is based first on the construction of analytical expressions describing the phenomena studied. Secondly (see section 7.3.3), we will examine the common elements of these approaches which will enable us to go back to the convergence between spatial dynamics and dynamic spaces.

7.3.1. *Computer models and simulation of spatial dynamics*

We have chosen to discuss the different concepts implemented in the simulation of a complex spatial system from the model of the ecosystem of the interior delta of the river Niger [KUP 99].

7.3.1.1. *An example: modeling of the ecosystem of the interior delta of the river Niger*

The interior delta of the river Niger, which is a vast inundatable zone with a million residents, produces approximately 10% of the domestic product of Mali. It constitutes a natural spasmodic and contrasting system and a social system structured to exploit its natural resources. The objective of modeling the delta is to address questions such as: how will this organized system evolve with the demographic pressure, the climatic changes, the technical evolution, the impact of hydraulic and agricultural developments? Or, how can we represent the system dynamics in its spatial extension so that we can take advantage of all of the elements in an integrated control perspective?

In this context, the objective is to highlight the interactions between phenomena of different natures, happening at different steps in time, over geographical spaces of different types. The idea is to approach and show the complexity corresponding to the large number of elements to integrate: different physical and biological sectors with the social, economic and political sectors, different resource system operations, different levels of spatial and temporal organization, different interests of the players in rural production, development, research and decision making.

The description of the elements of the ecosystem

The model of the interior delta of the river Niger presents a geographical space made up of stretches of river, channels, flood plains, fishery zones, agricultural sections, areas of pasture. In this space, fishermen, farmers, farmer-fishermen and herds operate. Throughout the seasons, the whole system evolves according to several processes:

– hydrological processes: rivers, channels and flood plains are emptied and filled up again. The flood mechanism is shown as a model of reservoir. The height of the water helps us to deduce the water surface of hydrological objects;

– hydrobiological processes: the reproduction of the fish occurs in flood plains. In each of these plains, it follows a logistical equation in which the carrying capacity depends on the flooded surface. Then, the diffusion of the fish follows a mixed equation depending on the connectivity of the system;

– vegetative processes: the crops and natural vegetation also follow their seasonal dynamics according to the state of the environment;

– development of resources and production processes: the fishermen and farmers move and distribute among themselves the fishery and agricultural zones to practice their activities. This distribution is determined by an allocation mechanism of labor market type. In this way, at each step in time, the need for labor is evaluated for each zone according to the seasonal activity and to the intensity of the flood. Labor availability is evaluated for each group and then the distribution is done;

– development of resources and production processes: the shepherds move across the pasture following their herd. It is the herd move model that is based on a similar mechanism of allocation: at each step in time, the carrying capacity of the herd is evaluated for each zone according to the flood and the herd already in place. If the number of cattle in place exceeds the carrying capacity, the herd will move to an accessible zone; if the move is impossible, then the sanitary state of the herd in place deteriorates; if the sanitary state of the herd deteriorates too much, animals are sold.

The processes are based on the relation between height and flooding, on the renewal and diffusion of fish, on the results of the state of the environment and of the actual decisions in fisheries, crops, pasture and, finally, on the balance between supply and demand of labor (decision making in fisheries and agriculture) and on the movement of herd when the environment is saturated (decision making in livestock farming). The driving force that controls all the mechanisms is the propagation of water in the network from a given flood height at the entrance of the delta. The other mechanisms involve the processes of renewal and sampling of natural resources and the processes of production economically speaking. All thematic sectors present in the model have in common mechanisms based on the different movements (of water, of producers, of bovines) within the delta, which technically and physically completes the geographic integration of the entire naturally inundatable region.

The main role of space

In the model of the interior delta of the river Niger, space occurs in two ways: i) as an element of description, the extent and the topology of the hydrological network that get modified, the economic players in movement; ii) as an element of explanation, through the effects of the maximum extent of flooded zones on the fish production, of the connectivity of the hydrological network on the fish diffusion in the hydrosystem, of time in relation to the traveling distances, of the carrying capacity limits determined by the variable extent of pastoral zones.

The interior delta as a natural region has been defined geographically by the fluvial flooding perimeter and functionally as an oriented, open network. The retained entities have a specific geographical meaning, since they are referenced by their exact location (plains, channels, connections) and their relative positions are respected. They have a socio-economic meaning, since they correspond to sectors of natural renewal, of sampling and of production by man of natural resources of the flooding area, and they also have a hydrological meaning. The notion of geographical space of that zone is used as an interface to understand, grasp and represent the processes of natural production of resources and the processes of their development. It is a good way to restore the complexity of the ecosystem and to understand the significance of each element in the group. The system's connectivity, or more precisely the hydrographical connections, the migration itineraries and fishing rights, constitutes an essential element. One of the consequences of this spatial choice is the restoration of the results in the form of localized data and maps.

In the model, it is not the strategies or the decision modes of economic agents that are is represented, but the behaviors – in terms of efficient practices and strategy results – and in particular spatial behaviors of homogenous groups of individuals (fishermen, farmers, herds). The model does not in fact present units of production, each having a revenue goal and management rules, of which it would deduce the choice of a zone to develop and a migratory behavior. It represents social groups residing in villages, each with a given behavior in terms of production and migration. The hypothesis is therefore that the spatial behavior of these groups is highly structured and that in the short term the induced structuring is stable and it can be quantified and serve as a basis for extrapolations[10].

The computer implementation of the model

It has resulted in a system structured in several layers – for example, the "hydrology" layer, the "hydrobiology" layer – following the example of geographical information systems, but this time considering the dynamics. Each of the layers groups a set of entities and a set of processes attached to it. Developed in this framework is an iterative process in which we focus consecutively on the whole (identification of the different layers and communication modes between layers) and the parts (definition and programming of entities and processes of each layer). The result, which is the integrated model, is constituted on the one hand of choices made in terms of communication modes between these layers and on the other hand of processes actually presented within these layers. An interface[11] enables the visualization of data, the production of simulations, and the visualization of the

---

[10] This is one of the reasons that led to retain a temporal extension of three years for simulations. In the longer term, this hypothesis becomes impossible: the changes of conditions obviously imply evolutions in behaviors.

[11] This can be viewed on http://www.orleans.fr.

results in the form of graphs or animations, the variation of parameters and the realization of sensitivity analyses. One of the expected products, apart from the formalization and inherent conceptualization of the simulation methodology, is the constitution of a platform of discussion and exchanges for a better coordination of the research in the different disciplines involved.

7.3.1.2. *The concepts of a computer model of spatial dynamics*

From the previous example, let us analyze a few important concepts of a computer model of spatial dynamics. The model of the interior delta of the river Niger is a big part of notions associated with the current vocabulary of environmental information systems [GAY 97]. It is built in fact from a representation of the world in terms of *entities*, which are linked by *relations* and characterized by *attributes*, and grouped in different information layers. It is important to come back to it by specifying how the model connects these notions to the consideration of space and that of the dynamics.

Let us first consider the concepts of entity, relation, attribute and their relation with space.

An entity is an abstraction of a piece of reality, conserving a certain instant character and close to the current or "naturalist" perception of things. It is an object to which observers attribute a persistence (which makes it possible to follow it through time) and for which they agree on a way to describe it. In our example, a flood plain geographically referenced in its exact place is an entity. A group of identified fishermen will also be an entity of which we will be able to follow the movements and the evolution of the income. In addition, an entity such as a river can be made up of other entities of different types such as the stretches of the river.

A relation is a characterization of links that exist within a set of identified entities. A distance between two fishery zones is a relation characterizing this pair of entities. The fact that a group of fishermen exploits a fishing zone, defines a relation between these two entities.

An attribute is an element of description of an entity or of a relation. We should remind our readers that the value of an attribute is not necessarily reduced to a number or a qualitative method. It can hold a more complex structure and be a vector, or even a continuous function. That is particularly true in the case of entities where we must describe the spatial registration or the evolution in time. We can then distinguish spatial attributes defining the spatial registration of an entity in a geometrical frame of reference (the coordinates of points of a contour) or through its spatial relations with other entities; non-spatial attributes such as biomass, specific composition of a river or the population of a village; finally, more complex attributes explaining the spatial variability of certain sizes: density of biomass at different areas of a river; variations of depth of a phreatic water; etc.

The relation with space is at the base of a distinction that we will discuss, in section 7.3.3 and that possesses several extensions. We can then find in the model of the interior delta two entity types:

– *space* entities: these are the entities whose registration in space is the element constituting their identity through their geographical location. That is the case with a flood plain, a channel, a village. Even if some of their spatial attributes evolve (form, spread), these entities remain attached to a portion of space, to the degree that they can be used to designate this portion of space (the name of the village thus designates a specific region);

– *player* entities: these are entities whose identity is defined without intrinsic link with their registration in space. On the timescale of the studied phenomenon, this identity must rely on other elements than the spatial registration *within the current time*, either because the entities can move, or because they can be born and die and therefore the portion of space they occupy, although set, cannot by itself define them. In the example of the interior delta, these are the groups of fishermen, farmers or herds. However, in other examples it could be trees in a forest.

We will now consider the dynamics themselves through the process, mechanism and control concepts.

In the same way that we define spatial attributes, we also identify temporal attributes indicating, in a time reference, the evolution of a variable (for example, the average height or of the flow of a river) or the start date of a specific activity. However, the static vocabulary of information systems (entities, relations, attributes) is not sufficient to build a model able to describe dynamics and enable us to experiment different scenarios in a simulator. The model of the interior delta calls for more flexible notions of *processes* and *mechanisms*.

In the model's vocabulary, a *process* represents a type of change occurring in an entity or in a group of entities and their relation. A process is characterized by its effect, an evolution of certain attributes and by the *mechanisms* which determine this effect. It is spatial if the effect modifies spatial attributes or if the mechanism involves such attributes. In this way, in the delta, some of the hydrological processes will have the effect of evolving the water height in flood planes and, as a result, the spatial extension of these entities. Their mechanisms develop water height in the channels and the topography of these plains at the same time. Distribution processes of agricultural zones between groups of farmers determine the allocation of such zones to which groups according to mechanisms resembling mechanisms of supply and demand adjustments on a market. An entity is always actively involved with at least one process, that of its own existence, which makes it move forward in time. In the view presented here, two entities are *interacting* when the state of one of the entities – the value of its attributes – or even its simple presence enter in the mechanism of a process affecting the state of the other entity or its existence. In this view, the processes are the ones interacting through the entities that they manipulate. The processes are the real driving forces of the dynamics.

Finally, the computer model of the interior delta of the river Niger, beyond the conceptual definition of the entities, of the processes and their effects and mechanisms, specifies a virtual computer world, i.e. a formalization where all the entities, relations and attributes are translated as computer objects, where all the processes, their effects and mechanisms are translated into algorithms and where all these objects and algorithms are inserted in an architecture and a control mode. This *architecture* and this *control* obviously occur in the implementation, but come from choices – for example, with rules behind the temporal sequence of the different algorithms – which can have heavy consequences in terms of the validity of simulations. They must be explicit and are therefore part of the model itself. It is mainly at this level that we start to talk about agents and to define those that ensure the control of the whole from the top of the hierarchy of controls. We will come back to this in section 7.4. At the level of modeling of spatial dynamics, we will only emphasize that, in the example represented, it is to space entities (hydrologic zones, agricultural zones, pasture zones) and not to player entities (the groups of farmers, of fishermen), that is attributed the largest part of the control of the system. This is a deliberate choice.

7.3.2. *Mathematical models of spatial dynamics*

7.3.2.1. *Eulerian and Lagrangian approaches*

Spatial dynamics, their mechanisms and the processes that control them have been the subject of generic models for a long time. The most traditional mathematic approach is based on the modeling of equations with partial derivatives (PDE). The world is seen as a continuous space. The state of the world, at each moment t, is described by a set of measurable quantities (weather, pressure, density, mass, concentration, etc.) defined at each point x. Values are associated with these quantities, real numbers, $e(x,t)$: these are the values of a numeric variable e (or a vector of variables if several quantities are involved). To study the spatio-temporal dynamics of the system means following and understanding the evolution of these values in space and time. The processes that explain the modifications of these values, i.e. the changes of state of the system, are modeled by equations describing the links between infinitesimal variations of the type:

$$\frac{\partial e(x,t)}{\partial t} \text{ and } \frac{\partial e(x,t)}{\partial x}$$

(partial derivatives), according to the limiting conditions of the studied area.

Let us take the example of very simple spatial dynamics, which are made up of particles moving in a one dimensional space, under the effect of mechanisms such as random and density-dependence movement (see Figure 7.1).

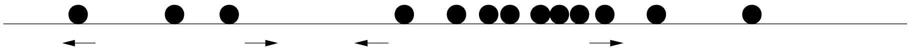


Figure 7.1. *Particles flee high density zones*

Consider the case where the particles are in large enough numbers so that we can consider, at each location x in space and at each moment t, a density $\rho(x,t)$ (a particular case of unidimensional variable e, $e = \rho$, and space that is also unidimensional, x scalar). The dynamics will result in processes that can be modeled by the following PDE:

$$\frac{\partial \rho(x,t)}{\partial t} = -\beta \frac{\partial \rho(x,t)}{\partial x} + \gamma \frac{\partial^2 \rho(x,t)}{\partial x^2} \qquad [7.1]$$

where coefficients β and γ represent, respectively, the intensity of the density dependent repulsion and a diffusion parameter characterizing random activity.

Generally, we can mathematically resolve the PDEs when the factors that they involve are simply expressed and for simple limiting conditions. We then have at our disposal, for the study of the dynamics that they model, analytical reference solutions for the many more complex cases where we must resort to numeric simulations. Spatial dynamics simulations are then carried out in a traditional way, using finite differences or finite element schemes, where space and time are represented by variable discrete space and time steps. The values of type $e(x,t)$ are averaged on the spatial reference cells.

In mathematics, as in physics or in quantitative biology, we qualify this type of model as Eulerian: whether the spatial support is continuous in the form of an infinite set of points or discrete in the form of a set of cells, it is at this support that the Eulerian approach links a variable, or a density in our current example.

We can generalize this vocabulary to a resolutely discrete approach, of the cellular automaton network type, where space is represented, at the very start of the modeling, as a finite set of spatial cells, and the variable studied $e(x,t)$ is replaced by a finite number of states $E_i(C_i,t)$. Figure 7.2 thus illustrates a unidimensional network where the state of each automaton gets the value of the integer number of particles contained in a cell.

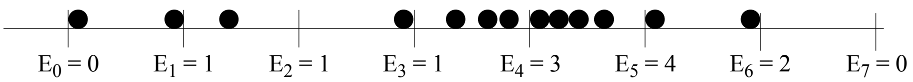


Figure 7.2. *A network of cellular robots to simulate density dependent movements*

The processes are then represented in the form of transition functions. The mechanism of "density-dependence" can be modeled in the form of a process modifying in an iterative way the state of a cell according to the state of its neighbors. Without going too much into detail[12], and without including the random activity, a transition function (operating on each automaton i from the states of its neighbors on the left $i - 1$ and on the right $i + 1$) can be defined from the following principle; for each spatial cell i and for each step in time:

$$[E_{i-1}, E_i, E_{i+1}] \rightarrow [E_{i-1}, E_i + \mathrm{sign}(E_{i+1} - E_i) + \mathrm{sign}(E_{i-1} - E_i), E_{i+1}] \qquad [7.2]$$

where $sign(y) = 1$ if $y > 0$ and -1 if $y < 0$[13].

However, mathematics has also focused on spatial dynamics through the Lagrangian approaches. We consider that the variables whose values evolve in space and time are relative to intrinsic quantities that can be defined independently of the spatial support.

The most simple example involves the current case where the modification of a variable of macroscopic state $e(x,t)$, defined from a fixed spatial support, has an equivalent in the form of an integration of the dynamics of basic entities which are

---

[12] The definition of a network of cellular automata representing particle flow processes is somewhat tricky because we must respect the constraints of conservation and of non-negativity of the number of particles in each portion of space [VAN 99].

[13] The triple $[E_{i-1}, E_i, E_{i+1}]$ represents the combined state of the three contiguous automata at the same moment t. The arrow indicates the way in which this triple is transformed when we move to the next time step. The indicated formula is exact in the case where the three states of the triplet are non-zero, and must be completed in the other cases.

defined at a finer scale and in movement in space. This is the case with the previous example where the density is associated with a number of particles of which we can define the existence independently from their position at time t and where the variation of density can be equivalent to the result of particle moves. A Lagrangian approach results then in the modeling of the movement of each particle considered individually (see Figure 7.3), which involves the same particles, but with a different point of view since we mainly consider spatial cells whose particle numbers have been modified.

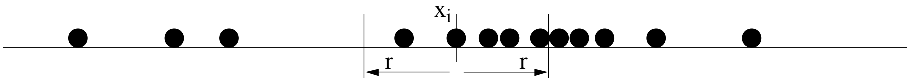


Figure 7.3. *A Lagrangian approach to simulate density dependent movements of individualized particles (compare with Figures 7.1 and 7.2)*

In our example [GRÜ 94], each particle p_i has, among other attributes, a position $x_i(p_i,t)$. We measure the velocity of each particle p_i according to the difference between the number of particles respectively located to its left and to its right in a radius r (see Figure 7.3).

A density dependent process then may be explained by a system of differential equations with, for each particle i:

$$\frac{dx_i}{dt} = -\beta \sum_j u(x_j - x_i) sign(x_j - x_i) \qquad [7.3]$$

where $u(x_j - x_i) = \begin{cases} 1 \text{ if } |x_j - x_i| \leq r \\ 0 \text{ if } |x_j - x_i| > r \end{cases}$ formalizes the notion of neighboring particles in the case of a set of moving particles.

The random activity mechanism can be added by introducing a random variable with zero mean ε_i in what becomes a system of stochastic differential equations:

$$\frac{dx_i}{dt} = -\beta \sum_j u(x_j - x_i) sign(x_j - x_i) + \gamma \varepsilon_i \qquad [7.4]$$

Several variations of the Lagrangian approaches enable us to consider references of space and time, continuous or discrete. We will talk, for example, of random walkers for a finite number of entities in movement on a regular cell grid, or of *particle-tracking* to model the movement of solute particles in a velocity field, as an alternative to PDEs based on concentration type variables

There are several studies focusing on the conditions for equivalence of Eulerian and Lagrangian formulations, with the help of methods borrowed from statistical physics (Boltzmann). This is why we can start with a Lagrangian type model to describe individual behaviors and then aggregate by spatial zones to obtain an Eulerian formulation that is easier to manage for large numbers of particles [GRÜ 94].

7.3.2.2. *An example on water runoff modeling*

Finally, let us give an example from one of our own application fields concerning the modeling of water runoff on a topographical surface (a soil that is assumed to be bare and impermeable). *Shallow water equations* are a simplified form of Navier-Stokes equations in fluid mechanics, when the depth of the water is small and we can namely neglect vertical accelerations. The main variable is the water height $h(x,t)$.

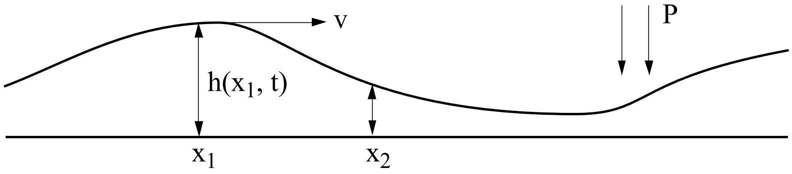


Figure 7.4. *An Eulerian approach to simulate modifications of water height*

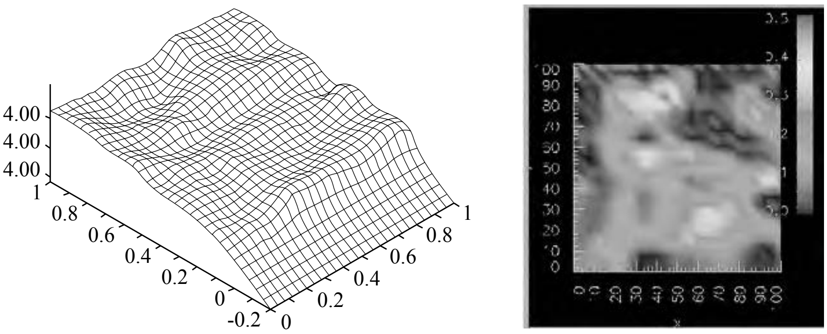
In the mono-dimensional case of Figure 7.4, with an external input of rain with constant intensity p, we can describe the following PDE:

$$\frac{\partial h(x,t)}{\partial t} = -\frac{\partial h(x,t)v(x,t)}{\partial x} + p \qquad [7.5]$$

The velocity v of water movements depends on the gradient of the water height and therefore on x (as opposed to β in equation [7.1] with dependence density). It also depends on the topographical gradient (let us note it is zero in Figure 7.4):

$$v = f\left(\frac{\partial(z + h(x,t))}{\delta x}\right) \qquad [7.6]$$

The numeric resolution of these equations on a regular spatial grid mainly enables us to simulate the evolution of the spatio-temporal field of water heights as indicated in Figure 7.5 (bidimensional version).



(a) (b)

Figure 7.5. *(a) The topography of an experimental parcel of 1 m^2; (b) simulated water height field during rain (clear zones show high depth of water and dark zones low depth of water) (according to [LEO 00])*

An approach that we have qualified as multi-agent, but that can also be viewed as a particle-based and Lagrangian version of the same phenomenon, has been developed by Servat [SER 00]. The water is viewed as an individualized entity population, agents or "water-balls" particles, brought by rain and in movement on the topographic surface represented as a continuous space (see Figures 7.5 and 7.6).

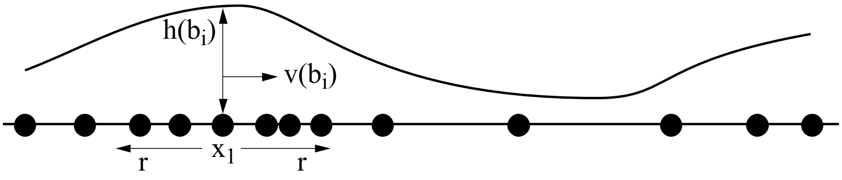


Figure 7.6. *Unidimensional representation of a continuous variable water height from abstract entities, individualized "water-balls" (compare with Figure 7.4)*

For each water-ball b_i at position $x_i(b_i,t)$ (see Figure 7.6), the idea is to calculate its velocity from information taken in a local environment defined by a surveyed neighborhood of radius r. The velocity is a function of the topographical gradient of this neighborhood and also of the position of the set of water-balls that it contains. It is the same function as that used in the PDE formulation ([7.6]) (although ignoring the accelerations for sake of simplicity sake), but the water height is estimated here according to the mass of neighboring water-balls, using a interpolation scheme[14] W weighted by distance to the central water-ball i involved:

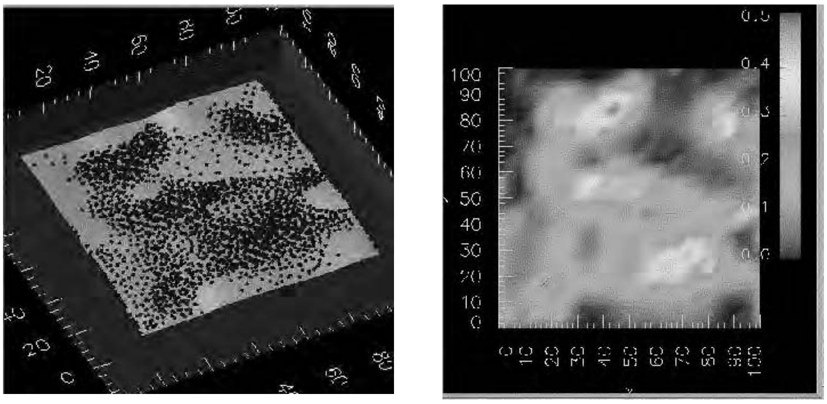
$$h(b_i) = \sum_j m_j W(x_j - x_i, r) u(x_j - x_i) \qquad [7.7]$$

where u has the same meaning as in equation [7.3].

For each particle i, we have:

$$\frac{dx_i}{dt} = v(x_i) \qquad [7.8]$$

where v is calculated by using equations [7.6] and [7.7].



(a) (b)

Figure 7.7. *(a) Moving water-balls. (b) Equivalent water height field to compare with Figure 7.5b (same data, same moment t) (according to [SER 00])*

The results of the simulations are very similar to those obtained by using PDEs (compare Figures 7.5b and 7.7b). Due to this Lagrangian approach, we can follow the trajectories of each water-ball, whereas it is only the history of water volumes

---

[14] We use a core function W defined on the neighborhood of radius r and using discrete sums of the SPH type [MON 92], which are also used for the calculation of gradients.

present on a given cell that could be deduced most easily from the approach by Eulerian simulation.

7.3.3. *Computer and mathematical models of spatial dynamics toward convergence*

The approach of computer modeling and the mathematical approach that we have presented with some examples seem to come from a different state of mind and, due to this they are often opposed [FER 95]. The mathematical approach addresses systems made up of multiple elements but with identical nature, on which a small number of rules or laws of evolution operate. The formalization of these elements and of these laws is part of a mathematical language – whose terms are variables, functions, operators and equations. This formalization constitutes a system "manually" processable by symbolic processes. Its primary vocation is to be a complete explanation support and a thinking support which makes it possible to analytically deduce some properties. On the contrary, the computer approach embraces a dynamics made up of a set of various heterogenous entities, processes and mechanisms described in a computer program. The potentialities of this program, for instance, the properties of the dynamics it represents, cannot be revealed or deduced without actually executing it. However, beyond these differences, a convergence holds our attention. In each of these two types of approaches the same duality on how to address the relation between space and dynamics appears.

7.3.3.1. *A common duality: Eulerian point of view and Lagrangian point of view*

The mathematical approaches have emphasized the distinction between Eulerian formulations where the equations specify what happens in each fixed spatial unit and Lagrangian formulations where the equations specify what happens in the elements. The approach illustrated in the example of the interior delta of the river Niger has considered two types of entities: space entities, which are intrinsically identified to a portion of space, and player entities, whose identity resides outside the spatial registration, which then becomes only a contingent attribute. The parallelism of these distinctions seems clear. It has a general range if we consider that several spatial dynamics can be analyzed as a result of *flows*: movements (transfers, migrations, transportation, dispersion), demographic processes (birth, death) modifying portions of space by including and deleting, various entities or substances (energy, fluid, matter, human beings, information).

The Eulerian point of view, in a broad sense, focuses on the evolution of phenomena from arbitrary fixed locations and follows a characterization of the locations over time[15]. By definition, it assumes the existence of a substrate of places, which can be either a set of points with reference coordinates, (for example, measured points) or, in a more general way, a set of space entities having a shape and inserted in a topology. It is concerned with knowing how these space entities are transformed in their shapes, properties and relations of different types (not only spatial) that they mutually maintain. It attempts to model the processes considering the studied phenomena at the entity level.

The Lagrangian point of view, also in a broad sense, considers a layer of reality made up of player entities that move in the physical space, meet and interact in multiple ways. It focuses directly on these entities, on their particular destiny, and in particular on their spatial trajectory and on the evolution of the vision of the world they have from their own reference point[16]. It uses them as analysis units and attempts to use the description of their behavior (movement in particular) as the bases of the explanation of the phenomena described at a higher organization level.

This coexistence of two viewpoints, one focusing particularly on space (space seen as a reality existing in itself) and the other on its content (space seen as a structure of relations, the only reality residing in the linked objects) is the subject of many debates in several disciplines, from physics to human geography. The Eulerian point of view is in harmony with the scientific movement that uses spatial structures as a starting point and attempts to dynamize them. The Lagrangian point of view is more in keeping the functional approaches and attempting to spatialize them. The history of these disciplines shows that it seems difficult to ignore one or the other point of view and that it might be more fruitful to organize their complementarity.

7.3.3.2. *Source and necessity of the comparison: simulation and its limits*

The computer models often give the impression of a certain opacity: their execution produces an object, i.e. the set of data constituting the result of a simulation, but does not generally provide by itself a way to represent a property of this object and *a fortiori* to prove it. A large portion of the research effort in complex systems modeling consists of fighting against this opacity: the choices made in the construction of the virtual world must be made more explicit and legible, we must create ways to be as certain as possible of the compliance of the result of the simulations with these explicit choices and even recognize and automatically signal the appearance of a property [SER 00]. The construction of formalisms capable of providing symbolic and easily manipulated expressions of computer models would enable us to consider the validation and the demonstration

---

[15] The Eulerian approach inspires the most current observation devices.

[16] The biographical surveys, the follow-up of markers, etc. are examples of observation devices of Lagrangian inspiration.

of properties in better conditions. This research direction is currently being explored. A few paths have been proposed: either from a potential comparison of existing analytical models with the studied processes or from formalisms that are purely computer-based [BAK 00, FER 00].

On the other hand, the properties within reach of an analytical reasoning from the mathematical statement of the initial conditions and laws of dynamics are often themselves in small number: although this statement might remain highly expressive, it will largely resist further analysis. We are then driven to search for ways to translate it into a computer simulation program if we want to go further. This program can also be seen as a virtual world, describable in terms of entities, processes, and control diagrams. As an example, a description in these terms of the density-dependence model seen in section 7.3.2 follows.

By adopting a Lagrangian point of view, the construction of a virtual world can involve:

– player entities: the particles, which are characterized by an attribute of position;

– spatial relations between the particles, relations of proximity *P: iPj* (*i* is considered similar to *j*), if the distance between particle *i* and particle *j* is lower than a certain parameter *r*;

– processes: the basic movements. Their effect consists of the modification of the position of a given particle *i* of finite quantity *dx*; following an algorithm enabling us to calculate *dx* from the positions of the particles that are near in the sense of *P* and of the chosen finite time step *dt*;

– a control scheme: for example, a synchronous scheme where each movement algorithm will calculate the position of a particle, at moment $t + dt$, from the position of this particle and of the particles that are near, such as they are at time *t*.

Such a computer model works in a manner that is compliant to the mathematical formulation, as long as the choice of a finite time step does not introduce any divergence, which must be proven. It also operates in an efficient way if the number of particles is small enough. In fact, the necessary update at each time step of proximity relations use a lot of computing time: in the execution of the movement algorithm of particle *i* we must review all the particle pairs (*i,j*) to determine *iPj*. This problem is not specific to spatial processes themselves: it happens each time a large number of relations evolve in the modeled dynamics. In the example presented as in many other similar cases, the method that we use to increase performances consists of completing the Lagrangian modeling viewpoint by a pragmatic Eulerian viewpoint. In this way we introduce additional space entities. These cells constitute the reference grid in which the particles move; we also introduce the relations (adjacency between cells, temporary affiliation of a particle to a cell), the processes and the modifications of control necessary for an efficient use of the system.

Using such a type of formalism of entities, relations and processes to describe a simulator of dynamics which could be mathematically described in a very simple manner can seem out of place. However, the objective was to show the possibility for computer models to include mathematical situations like particular cases, but to design them in a way that naturally enables the introduction of more complex behaviors.

7.4. The multi-agent approach in spatial dynamics modeling: a point of view

The spatial dynamics that we have presented result from a tangle of interactions. In order to conduct their analysis, we propose, from previous reflections, some concepts and an implementation methodology.

7.4.1. *The methodology*

Spatial dynamics modeling implies the development of a series of steps including:

– the identification, in the reality which is to be represented, of a finite set of entities and relation types;

– the characterization of these types of entities and relations by a finite number of attributes;

– the identification of the types of processes and the characterization of their effects and mechanisms in the form of algorithms;

– the definition of control schemes managing the temporal sequence of execution of the processes;

– the coding itself, trying to respect the defined structure with entities, processes and control;

– the creation of an “initial world” made up of instances of the types of entities, relations and processes;

– several executions of the program, producing different simulations based on different *scenarios*.

Research for the highest isomorphism between on the one hand the structure of all the entities, relations and processes affecting them and on the other hand the architecture of the computer program is greatly emphasized. This isomorphism research justifies that we consider the program as a “virtual world”. It favors the use of object programming languages, without excluding other programming languages. The program translates the entities, relations, algorithms and controls in a set of *computer objects*.

7.4.2. *Hierarchy of choices and the place of agents: an example*

In the current absence of a real “agent-oriented” language, the contribution of researchers on multi-agent systems is mainly done at the concept level. It is important here to clarify their place in spatial dynamics modeling.

We will use the same example that has enabled us to illustrate the passage from a mathematical model to a computer model, that of a “density dependent” dynamics. With the chosen Lagrangian viewpoint, the model is built around the behavior of particles. It has been proven useful for performance reasons to introduce secondary entities such as space intervals. When we reach the coding step, the entities become computer objects and we must ask ourselves in which objects the movement mechanism algorithms and their control will be programmed. Will it be the objects that represent the particles that will calculate their movement from information stored in the objects representing the internals, or will it be the objects representing the intervals that will calculate the movement of the particles and proceed to the necessary location transfers? Both choices are possible and are based on architecture and control questions already discussed. They define, on the one hand, which objects will be candidates to the function of agents, encapsulating the true driving forces of the dynamics and, on the other hand, which objects will remain reserved for storing information.

Once the choice is made, we need to determine which agent model to use. We may want to guide programming of these agents to specific structures indicating their internal organization and their mode of interaction. Such structures are partially imposed by the use of a platform. A platform restrains the programmer through its own concepts[17]. They can result from the deliberate choice by the programmer of an *agent metaphor* to which he will try to comply. Here again, we have a large range of metaphors at our disposal (autonomous robots, animals, human beings) leading to reactive agents interacting through an environment, or to agents that are cognitive and communicating with structured messages. Continuing on the model of “density-dependent” phenomena, and by assuming that we have

---

[17] We can mention here two general platforms: CORMAS, developed by CIRAD, presents to the user an interface enabling him to specify his multi-agent universe; SWARM, developed by the Santa Fe Institute, is presented as a library of programming structures which offers powerful tools for system inspection and simulation procedure.

chosen to designate particles as system agents, the programmer could also adopt the *robot metaphor* and consider particle agents as if they were robots moving in a real environment. The programming of the particle agent, if we want to use the metaphor to term, will need to structure the computer object that represents it in different elements that will play the role of sensors, effectors, deliberation modules, control and will encapsulate the corresponding processes (of perception, action, etc.). However, in this venture, we must not forget that, contrary to real robots, *the environment itself must be programmed*.

7.5. Conclusion

The viewpoint defended here is that the use of multi-agent concepts for computer modeling is a question of style, where the benefit results from the constraint that it exerts. Of course, this benefit is all the more ensured as the constraint is simply formulated and as its application is brought to term.

As shown in the example that we have developed, this constraint consists of, first, once the entities and processes of the model are identified and described, imposing explicit choices in the architecture of their computer translation and, in particular, in the system agents of computer objects that will encapsulate the necessary driving forces of dynamics. We recall that it is the principle of this constraint that is beneficial and that, in practice, the field of possibilities is wide open. We have been able to see that in terms of spatial dynamics, totally opposite choices can be legitimately adopted[18], whether we attribute the majority of the control and the agent status to objects representing space entities, as in the model of the interior delta of the river Niger, or to objects representing player entities, as in the numerous models where the agents represent individuals or social groups, or to objects not representing any entity but the researchers themselves (the “specialists”) with their knowledge and their formalizations of studied processes.

The constraint also consists of imposing the explanation of the choices in the internal agent structure and their modes of interaction once the agents are determined. Here again, several options are available: the advanced application of certain metaphors may sometimes seem as over complicated, but the goal is always to structure the programming, to make it as legible as possible and facilitate its evolution as the model becomes more complex.

A debate regularly shakes up the community of researchers using multi-agent simulations. We must determine if the choices that we qualify here of style only

---

[18] The work of Bousquet [BOUS 99] constitutes, in this regard, a particular example, since both choices are compared in the same problem.

involve the computer implementation of models, or if they are a more profound expression of a vision of the world that will determine the modeling itself. We feel it is necessary, as modelers to always distinguish the design from the implementation. Nevertheless, these two levels are not easily separated, especially since the implementation alone gives the final sense to computer modeling. It also seems that the intensive use of anthropomorphic metaphors can obscure the contribution of the models and lead to compare, sometimes in unproductive ways, computer modeling and mathematical modeling. We should, however, emphasize the fact that the use of these metaphors facilitates the presentation of models and their implementation, for the benefit of all. The solution to this dilemma would then reside in a cautious enthusiasm for the constant advances of the research on agents.

7.6. Bibliography

[AKN 98] AKNINE S., PINSON S., ZACKLAD M., "Un système d'agents récursilfs pour l'aide au travail coopératif", in *Systèmes multi-agents: de l'interaction à la socialité, JFIADSMA'98*, Barthès J.P., Chevrier V., Brassac C. (eds.), Hermes, p. 67-80, 1998.

[ASC 00] ASCHAN-LEYGONIE C., MATHIAN H., SANDERS L., MÄKILÄ K., "A spatial microsimulation of population dynamics in Southern France: a model integrating individual decisions and spatial constraints", in *Applications of Simulation to Social Sciences*, G. Ballot, G. Weisbuch (eds.), Hermes, p. 109-125, 2000.

[BAE 96] BAEIJS C., DEMAZEAU Y., "Les organisations dans les systèmes multi-agents", *4èmes journées du GDR PRC IA*, 1996.

[BAK 00] BAKAM I., KORDON F., LE PAGE C., BOUSQUET F., "Formalisation de modèles multi-agents par les réseaux de Petri: application à l'étude d'un système de gestion de la chasse à l'Est-Cameroun", *Actes du 5ème Colloque Africain sur la Recherche en Informatique*, *CARI 2000*, A. Corenthin, B. Philippe (eds.), INRIA, p. 265-272, 2000.

[BAN 05] BANOS A., CHARDONNEL S., LANG Ch., MARILLEAU N., THÉVENIN T., "Simulating the swarming city: a MAS approach", *The Ninth Int. Conf. on Computers in Urban Planning and Urban Management*, CUPUM 2005, London, 2005.

[BAT 96] BATARD E., "L'agent comme signe", in *IA Distribuée et Systèmes Multi-Agents*, J.P. Müller, J. Quinqueton (eds.), Hermes, Paris, p. 215-227, 1996.

[BEN 98] BENENSON I., "Multi-Agent Simulations of residential dynamics in the City", *Comput. Envir. and Urban Systems*, 22, p. 25-42, 1998.

[BOU 95] BOURGINE P., Models of Autonomous Agents and of their Coevolutionary Interactions, CREA Reports and Documents, no. 9517, 1995.

[BOUR 92] BOURON T., Structure de communication et d'organisation pour la coopération dans un univers multi-agents, PhD Thesis, Paris University 6, 1992.

[BOUS 93] BOUSQUET F., CAMBIER C., MULLON C., MORAND P., "Simulating the interaction between a society and a renewable resource", *Journal of Biological Systems*, 1(1), p. 199-214, 1993.

[BOUS 99] BOUSQUET F., GAUTHIER D., "Comparaison de deux approches de modélisation des dynamiques spatiales par simulation multi-agents: les approches spatiales et acteurs", *Revue Européenne de géographie Cybergeo*, 1999.

[BOUS 01] BOUSQUET F., LE PAGE C., BAKAM I., TAKFORYAN A., "A spatially explicit individual-based model of blue duikers population dynamics: multi-agent simulation of bushmeat hunting in an eastern Cameroonian village", *Ecological Modelling*, 138(1-3): 331-346, 2001.

[BRA 95] BRASSAC C., PESTY S., "Cooperation in Multi-Agent Systems: Behaviour or Conduct?", in *DIMAS 95 Decentralizes Intelligent and Multi-Agent Systems*, Krakow, Poland, p. 84-92, 1995.

[BRE 00] BRETON L., ZUCKER J.-D., CLEMENT E., "GranuLab: un laboratoire virtuel d'expérimentations pour la découverte scientifique en physique granulaire", in *Proceedings of RFIA'2000*, November 1999, Paris, France, p. 599-606, 2000.

[BRO 86] BROOKS R., CONNELL J.H., "Asynchronous Distributed Control System for a Mobile Robot", *SPIEE vol. 727, Mobile Robots*, Cambridge, MA, p. 77-84, 1986.

[BUR 96] BURA S., GUERIN-PACE F., MATHIAN H., PUMAIN D., SANDERS L., "Multi-agents systems and the dynamics of a settlement system", *Geographical Analysis*, 28, p. 161-178, 1996.

[COL 96] COLLINOT A., PLOIX L., DROGOUL A., "Application de la méthode Cassiopée à l'organisation d'une équipe de robots", in *Actes des 4^èmes^ journées francophones d'intelligence Artficielle et Systèmes Multi-Agents*, p. 137-152, 1996.

[DON 06] DONIEC A., ESPIÉ S., MANDIAU R., PIECHOWIAK S., "Non-normative behaviour in multi-agent systems: some experiments in traffic simulation", *IAT*, p 30-36, 2006.

[DOR 98] DORAN J., "Social Simulation, Agents and Artificial Societies", in *ICMAS'98 Proceedings of the 3rd International Conference on Multi-Agent Systems*, La Villette, Paris, Y. Demazeau (ed.), IEEE Computer Society, p. 4-5, 1998.

[DRO 95] DROGOUL A., CORBARA B., FRESNEAU D., "Manta: new experimental results on the emergence of (artificial) ant societies", in *The Computer Simulation of Social Life*, N. Gilbert, R. Conte (eds.), UCL Press, London, 1995.

[DRO 97] DROGOUL A., COLLINOT C., "Entre réductionisme méthodologique et stratégie intentionnelle, l'éthologie, un modèle alternatif pour l'IAD", in *JFIADSMA'97 Intelligence Artificielle et Systèmes Multi-Agents*, Quinqueton J., Thomas M.C., Trousse B. (eds.), Hermes, Paris, p. 307-322, 1997.

[DUP 92] DUPUY J.P., *Introduction aux sciences sociales: logique des phénomènes collectifs*, Ellipses, Paris, 1992.

[DUR 89] DURFEE E.H., LESSER V.R., CORKILL D.D., "Trends in cooperative distributed problem solving", *IEEE Trans. Knowl. Data Eng.*, KOE-11(1), p. 63-83, 1989.

[FER 95] FERBER J., *Les systèmes multi-agents. Vers une intelligence collective*, InterEditions, Paris, 1995.

[FER 96] FERBER J., "Coopération, intention et interaction. Une réponse à la pelouse fourmillière de Brassac et Pesty", in *IA Distribuée et Systèmes Multi-Agents*, Müller J.P., Quinqueton J. (eds.), Hermes, Paris, p. 273-280, 1996.

[FER 98] FERBER J., GUTKNECHT O., "A meta-model for the analysis and design of organizations in multi-agent systems", in *ICMAS'98 Proceedings of the 3rd International Conference on Multi-Agent Systems*, La Villette, Paris, Y. Demazeau (ed.), IEEE Computer Society, p. 128-135, 1998.

[FER 00] FERBER J., GUTKNECHT O., "Pour une sémantique opérationnelle des systèmes multi-agents", in *Systèmes multi-agents. Méthodologie, technologie et expériences*, Pesty S., Sayettat-Fau C. (eds.), Hermes, Paris, p. 39-55, 2000.

[FIA 98] FIANYO E., TREUIL J.P., PERRIER E., DEMAZEAU E., "Multi-Agent Architecture Integrating Heterogenous Models of Dynamical Processes: The Representation of Time", in *Multi-Agent Systems and Agent-Based Simulation*, Sichman, Conte, Gilbert (eds.), Springer-Verlag, LNAI series no. 1534, Berlin, p. 226-236, 1998.

[GAY 97] GAYTE O., LIBOUREL T., CHEYLAN J.P., LARDON S., *Conception de Systèmes d'information sur l'environnement*, Hermes, Paris, 1997.

[GRI 99] GRIMM V., "Ten years of individual-based modelling in ecology: what have we learned and what could be learned in the future", *Ecological Modelling*, 115, p. 129-148, 1999.

[GRÜ 94] GRÜNBAUM D., OKUBO A., "Modeling social animals aggregations", in *Frontiers of Theoretical Biology*, S. Levin (ed.), Springer-Verlag, Lect. Notes in Biomathematics no. 100, p. 296-321, 1994.

[GUT 98] GUTKNECHT O., FERBER J., LIEURAIN E., "Des modèles hétérogènes de simulation par systèmes multi-agents", in *Modèles et systèmes multi-agents pour la gestion de l'environnement et des territoires*, *Actes du colloque Smaget*, Cemagref Editions, p. 433-444, 1998.

[HEM 04] HEMELRIJK C.K., "The use of artificial-life models for the study of social organization", in *Macaque Societies: A Model for the Study of Social Organization,* B. Thierry, M. Singh, W. Kaumanns (eds.), Cambridge: Cambridge University Press, pp. 295-313, 2004.

[HEM 05] HEMELRIJK C.K., WANTIA J., GYGAX L., "The construction of dominance order: comparing performance of five methods using an individual-based model", *Behaviour*, 142, 1037-1058, 2005.

[HOC 98] HOCKNEY R.W., EASTWOOD J.W., *Computer Simulation Using Particles*, Institute of Physics Publishing, London, 1998.

[JEN 93] JENNINGS N.R., "Commitments and conventions: the foundation of coordination in multi-agents systems", *Knowl. Eng. Rev.*, 8(3), p. 223-250, 1993.

[KIS 92] KISS G.R., “Decoupling agents from their environment: combining symbolic and reactive architectures”, in *Decentralized Artificial Intelligence*, E. Werner, Y. Demazeau (ed.), Elsevier/North Holland, Amsterdam, p. 115-127, 1992.

[KRE 01] KREFT J.-U., PICIOREANU C., WIMPENNY J., Van Loosdrecht M., “Individual-based modelling of biofilms”, *Microbiology*, 147, 2897–2912, 2001.

[KUP 99] KUPPER M., “Modélisation intégrée d’un écosystème inondé et gestion des eaux: le cas du delta intérieur du Niger au Mali”, in *Actes des journées scientifiques FRIEND-AOC-ZTH-AAH*, Yaoundé, November 30-December 2, UNESCO Editor, 1999.

[LAM 02] LAMJED B., “Agent-based interaction analysis of consumer behaviour”, *Proceedings of the International Conference on Autonomous Agents*, Bologna, Italy, p. 184-190, 2004.

[LAR 98] LARDON S., BARON C., BOMMEL P., BOUSQUET F., LE PAGE C., LIFRAN R., MONESTIEZ P., “Modéliser les configurations et les stratégies spatiales dans un système multi-agents pour la maîtrise de dynamiques d’embroussaillement”, in *Modèles et systèmes multi-agents pour la gestion de l’environnement et des territoires*, *Actes du colloque Smaget*, Cemagref Editions, p. 169-187, 1998.

[LEN 94] LENAY C., “Intelligence artificielle distribuée: modèle ou métaphore des phénomènes sociaux”, *Revue Internationale de Systémique*, vol. 8, no. 1, p. 1-11, 1994.

[LEN 95] LESNE A., *Méthodes de renormalisation, phénomènes critiques, chaos, fractales*, Eyrolles, Paris, 1995.

[LEP 96] LE PAGE C., CURY P., “How spatial heterogenity influences population dynamics: Simulations”, *SeaLab, Adaptive Behavior*, 4, (3/4), p. 249-274, 1996.

[MAR 90] MARUICHI T., ICHIKAWA M., TOKORO M., “Modeling Autonomous Agents and Their Groups”, in *Decentralized Artificial Intelligence*, Y. Demazeau, J.P. Müller (eds.), North Holland, p. 215-234, 1990.

[MAT 00] MATHEVEY R., Usage des zones humides camarguaises: enjeux et dynamique des interactions environnement-usagers-territoires, Jean Moulin-Lyon III University Thesis, 2000.

[MES 94] MESLE R., ICHTYUS: architecture d’un système multi-agents pour l’étude des structures agrégatives, DEA, LAFORIA report, Paris University 6, 1994.

[MES 06] MESKALI M., BARBET I., ESPIÉ S., BOOTSMA R., “Perceptual and motor strategies of car drivers in regulating their speed of approach to a preceding vehicle”, *Le Travail Humain*, vol. 69, no. 2, 183-207, 2006.

[MIL 99] MILLISCHER L, GASCUEL D., “Hétérogénéité spatiale et tactiques de pêche. Approche des comportements de pêche à l’aide d’un modèle individu-centré”, in *Les Espaces de l’halieutique*, *4ème Forum Halieumétrique*, Rennes 29 June-1 July 1999.

[MON 92] MONAGHAN J., “Smoothed particle hydrodynamics”, *Annu. Rev. Astron. Astrophys.*, 30, p. 543-574, 1992.

[OCC 97] OCCELLO M., DEMAZEAU Y., "Vers une approche de conception et de description récursive en univers multi-agents", in *Actes des 5èmes journées francophones d'intelligence Artficielle et Systèmes Multi-Agents, JFIADSMA'97*, Quinqueton J., Thomas M.C., Trousse B. (eds.), p. 143-158, 1997.

[PIC 91] PICHON G., MULLON C., REMY D., "Simulation du cycle des filaires", *Colloque L'utilisation de l'informatique dans l'enseignement de la biologie et de la géologie*, ENS/CNRP, 1991.

[PIC 04] PICIOREANU C., KREFT J.-U., VAN LOOSDRECHT M., "Particle-based multidimensional multispecies biofilm mode", *Applied and Environmental Microbiology,* vol. 70, p. 3024-3040, 2004.

[ROU 98] ROUCHIER J., BARRETEAU O., BOUSQUET F., PROTON H., "Evolution and Co-evolution of Individuals and Groups in Environment", in *Proceedings of the 3rd International Conference on Multi-Agent Systems, ICMAS'98, La Villette, Paris*, Y. Demazeau (ed.), IEEE Computer Society, p. 254-260, 1998.

[SAN 06] SANDERS L., FAVARO J-M., GLISSE B., MATHIAN H., PUMAIN D., "Artificial intelligence and collective agents: the EUROSIM model", *Cybergeo*, 2006.

[SCH 97] SCHWEITZER F., "Active Brownian Particles: artificial agents in physics. Stochastic dynamics", *LNPI 484*, p. 358-371, 1997.

[SER 00] SERVAT D., "Distribution du contrôle de l'action et de l'espace dans les simulations multi-agents de processus physiques", in *Systèmes multi-agents. Méthodologie, technologie et expériences*, S. Pesty, C. Sayettat-Fau (eds.), Hermes, Paris, p. 343-356, 2000.

[SHI 98] SHIN Y.J., CURY P., "OSMOSE, a multispecies individual-based model to explore the functional role of biodiversity in marine ecosystems", in *Ecossystem Considerations in Fisheries Management, Lowell Wakefield Fisheries Symposium*, University of Alaska Sea Grant College Program, 1998.

[SHO 93] SHOAM Y., "Agent-Oriented programming. Artificial Intelligence", 60(1), Elsevier Science, p. 51-92, 1993.

[TEL 98] TELES V., DE MARSILY G., PERRIER E., "Sur une nouvelle approche de modélisation de la mise en place des sédiments dans une plaine alluviale pour en représenter l'hétérogénéité", *Compte rendu de l'Académie des sciences de Paris*, Sciences de la terre et des planètes, 327, p. 597-606, 1998.

[THO 98] THOMAS J.D., SYCARA K., "Heterogeneity, Stability, and Efficiency in Distributed Systems", in *Proceedings of the Third International Conference on Multi-Agent Systems, ICMAS'98*, La Villette, Paris, Y. Demazeau (ed.), IEEE Computer Society, p. 293-300, 1998.

[VAN 88] VAN IMHOFF E., POST W., "Méthodes de micro-simulation pour les projections de population", *Population*, no. 4, p. 889-932, 1988.

[VAN 00] VANBERGUE D., TREUIL J.P., DROGOUL A., "Modelling urban phenomena with cellular automata", in *Applications of Simulation to Social Sciences*, G. Ballot, G. Weisbuch (eds.), Hermes, p. 109-125, 2000.

[WER 96] WERNER E., “Logical Foundations of Distributed Artificial Intelligence”, in *Foundations of Distributed Artificial Intelligence*, G.M.P. O’Hare, N.R. Jennings (eds.), John Wiley & Sons, New York, p. 57-117, 1996.

[WIL 92] WILSON K., “Les phénomènes physiques et les échelles de longueur”, in *L’Ordre du Chaos*, Belin, Bibliothèque pour la science, p. 88-115, 1992.

[WOO 95] WOOLDRIDGE M., JENNINGS N., “Intelligent agents: theory and practice”, *Knowledge engineering review*, Cambridge University Press, 10(2), 1995.

[YER 06] YERGOU B., GARY W., “A simulation model of biofilms with autonomous cells: analysis of a two-dimensional version”, *Physica A* 362, Elsevier, p 382-402, 2006.

Chapter 8

From Image to Model: Remote Sensing and Urban Modeling

8.1. Introduction

The objective of this chapter is to show how the connection of several models from different formalisms and perceptions can be used in the research on the evolution of the organization of space. From simple, iconic models to mathematical models, all use observed reality as their reference and offer a schematic representation. In the following pages, we have attempted to link models of a different nature and for which the term “schematic representation” takes a variety of meanings. A first model uses the principles of signal physics to extract information on land surface from a satellite image and a second model corresponds to a demographic logic and is based on the representation of the characteristics and behaviors of urban populations taken from surveys. The satellite image is then used to serve as a foundation for a new interpretation of the potential model in the hope of determining the most attractive urban areas in a migratory perspective. Finally, the results obtained will be introduced in a dynamic model of residential mobility and of territorial reorganization. The methodology adopted is based on the connection of a group of “models” different in their objectives and in their formalization. Although conceptual differences exist between each step, the satellite image – or a direct derivative – constitutes the lead in this chapter. We will go in these sections from an iconic model to a spatial distribution model of urban forms reflecting the distribution of the population, to a spatial interaction model between locations and finally to the dynamic model of a complex urban system.

Chapter written by Françoise DUREAU and Christiane WEBER.

8.1.1. *A modeling of urban reality*

The comprehension of space by a satellite image is done through a model based on the physical realities of the elements[1] of the land's surface. The satellite information corresponds to an "elaborate schematic representation based on defined objectives" according to Haggett [HAG 73]. The image must be revealing of situations, subject to promote questioning and attract other views. In this sense, the degree of realism from the resulting product facilitates visual perception of the image and favors its interpretation through a schematization process. In this way, the use of a classification to interpret the satellite image's representation facilitates the viewing and understanding of agricultural and urban forms. Understanding the different spatial, cultural and technical contexts is essential to the perception of reality. It helps to separate structures within an image, to identify objects (blocks of houses, networks, green spaces, etc.) and to establish topological and neighborly relations. In the case of the urban image and beyond the recognition of visual elements, it is interesting to detect geographical elements, urban forms that we can associate with production processes of space and socio-spatial organization. "Evocative of their genesis and enhanced by the image, urban forms become witnesses to the action from the societies creating them, using them and interpreting them" [WEB 95, p. 85]. The formalization of reality from such an "iconic modeling" can suggest various approaches using the image for its radiometric characteristics, as well as for its semantic content and its information contribution at different observation levels. These possibilities will be addressed here so that we may highlight the fact that the satellite image is an element in its own right of our modes of comprehension of urban complexity. It locates urban forms, facilitates their identification, demonstrates dynamic reality and it also develops a modeling that is either explanatory of identified processes or predictive of possible evolutions.

The arrival of satellites with high and very high spatial resolution has become a major point of interest in the last 20 years for communities[2] and for social science researchers. The first satellite images used were mainly for the representation of a specific aspect of space that went beyond traditional spatial information (cartographic or from aerial photographs). We could make inventories or geographical reports faster and more complete. Later, the improvements done to spatial resolution enabled us to address not only the location of objects, but also

---

[1] Each element of the land's surface, a building, a tree or a river has physical properties that enable the remote sensing signal to differentiate them. Remote sensing is based on this phenomenon by sensing the energy sent by the elements when they are activated by the sun or any other energy source.

[2] According to Jensen [JEN 99], millions of dollars are spent each year on spatial information (aerial photographs or satellite images) to extract urban information that cannot be measured *in situ*.

their description, identification and integration in predictive reasoning, among others.

The relations between images and socio-economic characteristics have quickly been the subject of studies in terms of classifications of land use (based on the way the land is used) or land cover (associated with the biophysical characteristics of the elements), document updates as in transport networks, or even their traffic [JEN 98]. Currently, the demographic estimates [DUR 89, FOR 85, HAA 97, LIN 85, LO 86] or the quality of life analyses [AVE 93, CUR 97] constitute the new axes of satellite image use in urban studies, side by side with modeling of evolution dynamics [BAT 95, CLA 96, MEA 90], mainly due to the improvement of spatial resolution of the Earth's observation sensors.

8.1.2. *Objectives of the chapter*

The methodology presented here is didactic. Beyond the case study of Bogota, it combines, in one location, problems as varied as:

– residential mobility and its interactions with the socio-spatial organization of a city;

– modes of urban space transformation through densification and expansion dynamics;

– the integration between these two urban dynamics within spatial and temporal dimensions;

The question of production of information and understanding of these domains enabled us to connect different types of modeling with each conceptual field and methodological approaches which are inherent to it. This connection is seldom common and requires necessary efforts of formalization of geographical objects (natural, technical and anthropogenic) used, as well as the possible (re)use of the results at different levels. Our goal is to focus on presenting a reasoning approach rather than results already acquired and validated.

This chapter focuses on the socio-demographic study of the city of Bogota. The objective is to use this case study to show that satellite imagery is a source of interesting and original spatial information in a process of construction of knowledge and modeling of the dynamics of intra-urban space. The satellite image makes it possible to implicitly take into account a space about which it provides the characteristics, the georeferencing and the integration of results within information systems.

Figure 8.1 illustrates the procedure of the proposed methodology for this chapter. It includes four sections. The satellite image plays a different part at each modeling step, from a representation of reality (see section 8.1) to its use within the constitution of a survey base (see section 8.2) from which mobility surveys have been developed. With the results of image processing, we have also been able to emphasize the sensitive zones corresponding to the areas of probable development for the city of Bogota (see section 8.3). These labeled surfaces can be introduced in an investigation mode enabling the simulation of urban dynamics (see section 8.4).

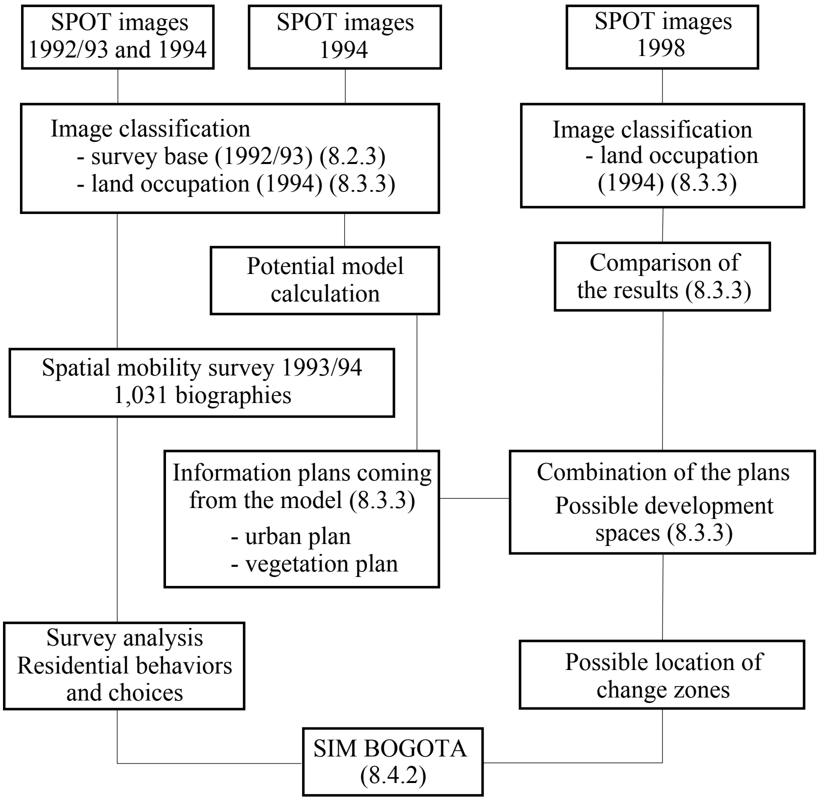


Figure 8.1. *The applied methodology for the analysis of the dynamics of Bogota*

The integration of these models leads to a final rich and productive sequence, where each step can be autonomous and still be enhanced by the previous one. The ever present urban space is formalized in a different way at each step: biophysical space, density of construction space, mobility space; or probable development space. The conceptualization of space progressively evolves, gradually showing or not, at each step a generally complete group of its environmental, social and spatio-temporal characteristics.

Source	**Date**	**Spatial coverage**	**Units**	**Information extracted**
Images: - SPOT XS* - Panchromatic - XS and Panchromatic	1992 1993 1994 and 1998	Bogota and periphery	10 m and 20 m pixels	- 1992/93: Survey base and stratification of mobility survey - 1994: Land occupation (initialization potential model) - 1998: Land occupation (verification of results potential model)
Aerial photographs	1992 1997	Bogota	Spatial resolution	Field-reality: initialization of image classification 1994
Mobility survey ORSTOM-CEDE (demo-statistical and anthropological)	1993 1994	11 districts of Bogota and its periphery	- Sector - Housing - Household - Person	- State of housing and population inventory in 1993 and 1994 - Population and housing flow between 1993 and 1994 - Migratory, professional, family biographies (retrospective: 1900-1993)
Census: - Housing - Population	1993	Bogota and its periphery	- Sector - Housing - Household - Person	- State of housing and population inventory - Flow of intermunicipality immigrants (over a lifetime and over 5 years) - Stratification of the mobility survey

Table 8.1. *The information used (* XS multispectral sensor)*

The choice for the experimentation site answers two imperatives: to show the adaptability of satellite information with these various objectives and the connection of the retained modeling forms. The city of Bogota (Colombia) was chosen because it is representative of the evolution dynamics of the major cities (see Box 8.1). The site offers optimal experimentation conditions due to the characteristics of the processes involved and of the elaborate corpus of information and knowledge available on the city. This corpus combines individual census data from several different years (1973, 1985, 1993), data from biographical surveys on the mobility trajectories of a sample of households (1,000 households surveyed in 1993) associated with family and professional contexts and, finally, information coming from the processing of satellite images (see Table 8.1).

Box 8.1. *The spatial and demographic dynamics of Bogota [DUR 99b]*

Moderate until the 1940s, Bogota's growth rate then starts to accelerate, and for a quarter of a century the growth rate of the population increases by over 6% annually and the spatial expansion grows at an even higher rate. At that time, the natural growth rate is combined with an intense migration: Colombia experiences during that time a large rural depopulation process polarized by the four largest cities in the country. In 1973, Bogota has nine times more residents (approximately 3 million residents) than in 1938, and is 12 times bigger. During the middle of the century, major spatial organization structures of Bogota are put in place and will only increase during the next decades: a very pronounced socio-spatial segregation opposing a rich North with a poor South where the illegal self-help housing largely dominates and a strong employment concentration along a center-north tertiary axis with a center-west industrial axis. From the 1970s, with the combined effect of the demographic transition and the decrease of intensity of the migratory flows, Bogota enters a period of slower growth (down to 3% per year) which integrates a growing number of peripheral towns. The previous expansion along the axes of circulation, that had left many interstitial spaces unoccupied now changes to a more general occupation of the space inside the urban perimeter and the start of vertical construction leads to the appearance of high density sectors, creating very high heterogenities in the occupation of the metropolitan space.

The distances created by the spatial expansion of previous decades and emphasized by transportation problems have, in the last 10 years, led to noticeable changes in the logic of the population's residential location. The competition for access to land increases: the city's spread must face the reliefs that border the town to the south and east, as well as the development of western farmlands now dedicated to intensive flower growing. In a very permissive regulatory frame, these factors act simultaneously and have recently been making profound changes in the metropolitan configuration. A dispersion of popular neighborhoods and a residential deconcentration of higher income populations are done at the same time as a rapid densification of certain surrounding spaces. In a growing shortage of urbanizable land, the global diagram of socio-spatial segregation inherited from a continuous dynamics over several decades gets more complicated with the emergence of new proximities between social groups. At the end of 1993, Bogota (including the metropolitan periphery) has over 6 million residents and the growth continues at a rate of 3.5% annually. The youth of the Bogota population, which maintains a high natural growth rate, is combined with very large gaps in demographic behaviors directly representing the large economic disparities of the population. While residing in the capital, which has the best health indicators of the country, the popular layers of Bogota experience a fertility rate similar to that observed in the Colombian countryside and suffer very high adult male death rate, which is linked to homicides and accidental deaths.

The chapter is divided into four parts. The first one introduces the use of image for socio-demographic information production (see section 8.2). The second introduces satellite image in a spatial interaction model (see section 8.3). The third part analyzes possible uses for the results of the two previous parts in dynamic modeling (see section 8.4) and a last section is devoted to discussion and a few conclusions.

8.2. The satellite image in the demographic information production

Representative of the dynamics of numerous major developing cities as seen in Box 8.1, Bogota shows the general characteristics that the observation of a large southern city must integrate:

– speed and flexibility in the production of information, imposed by the growth rates and the frequent and abrupt changes;

– the consideration of the whole agglomeration, which exceeds the urban perimeter of the district and spreads to peripheral municipalities;

– social and spatial disaggregation of the information in order to fully understand the important differences inherent to the city and satisfy management's needs;

– production of information relative to inventories (of urban space, housing, population) and to their transformation, and of information relative to flows (housing circulation, spatial mobility of intra and extra-urban populations).

In developing countries, urban management and research have long focused on the most visible manifestation of urban processes: demographic growth linked to a massive rural exodus and, what comes with it, peripheral expansion of the cities. Understanding the contemporary dynamics of these countries' major cities requires that we consider not only extra-urban migration and natural growth, but also intra-urban residential mobility. After several decades of fast growth, it has become the most important housing supply factor and consequently the main transformation factor of metropolitan systems in many major southern cities.

Satellite image could play a very important part in facing the limits of traditional demographic information production techniques, which are not very well adapted to the characteristics that the contemporary dynamics requires. The existence of relations between the forms of space occupation and the characteristics of residents opens the "social" field of applications to spatial remote sensing. The information provided by the satellite can be used as an observation vector in an information production process on city populations.

8.2.1. *The different phases of information production from satellite imagery*

Satellite imagery can be used in different ways in the production process of information on urban populations [DUR 95, p. 183-186].

In the first place, in the mapping phase that systematically precedes the general population census collection phase, imagery can be used to delimit urban zones, locate the zones that have witnessed change since the last available mapping (where we can see the necessity of devoting more resources to land mapping work) and, in certain types of urban fabric, identify and map blocks.

Subsequently, the satellite image can contribute to the improvement of existing information. With the help of the provided information on the morphology of a city, satellite imagery can spatially distribute demographic or social-economic data from administrative or census information, produced by coarse geographical units with no significance or no compatibility. The easiest example involves the measure of population density: locating non built-up zones on a satellite image and measuring the density by only considering the built-up zones, which could be inhabited, is clearly a better indicator of the reality of spatial distribution of the population of a city than the calculations done only on administrative sections. The satellite image can also be essential in addressing the question of geographical map units, where the diversity often constitutes an obstacle when we consider combining information of various sources (each organism diffusing information often has its own spatial sectioning corresponding to its specific needs) or information from one organism at different times [WEB 95].

Finally, satellite information can be used to produce socio-demographic information from area sampling. In fact, satellite images can be developed to produce intermediary information, which constitutes a step in an information production process involving sources other than remote sensing. The information researched is not part of the image, but the existence of links between the phenomenon that we wish to observe and the one present on the image facilitates and accelerates the production of information on this phenomenon. We can then take advantage of the comprehensive and finely located information on the occupation of the space provided by satellite images to quickly produce surveyed demographic or socio-economic information. The main goal is to use the satellite image as a survey base and develop the information on urban morphology provided to stratify a plan of area sampling that enables the selection of a survey sample for demographic or socio-economic purposes. This goal is not new; the development of area sampling methods on satellite images has directly benefited from the accumulated experience of the first applications on aerial photography in the 1950s in the USA [GRE 56, WAT 85] and later in developing countries [DES 93, IAU 83, VER 78].

8.2.2. *Area sampling method on satellite image: general principles*

In the absence of a current census base support, a team from IRD (Institut de Recherche pour le Développement) has developed an area sampling method on satellite image at the end of the 1980s. We will recall some of the general principles here[3]. It is an area sampling stratified in two degrees: in the first degree, we select geographic areas, which are in this case blocks; in the second degree, we select the surveyed units, which are in this case households, by a systematic equiprobable draw from the lists of households from each block in the sample. The satellite image will be used for:

– definition of the sample base: the urban limit is traced on the image from the information provided by it;

– stratification of the sample base[4]: an efficient and operational technique can be to stratify depending on the density criteria of developed areas;

– selection of a sample of blocks geographically spread throughout the city: we proceed to a systematic spatial draw from a points grid positioned over the image.

The first application, in 1987 in Quito (Ecuador), confirmed the interest of the method and we have been able to evaluate the costs of its implementation and the precision of the results. The sample that was used for the demographic survey had been selected in a rigorous way from the SPOT image. The 1993 selection of the sample for our survey on spatial mobility in Bogota highlights the application methods of this technique in a major city of several million residents with a SPOT image and exogenous information.

With this method, since then applied to other urban sites in different regions of developing countries [DUR 97, POI 97], it is now possible to quickly implement a sample survey in a city that does not have a traditional sample base and to produce demographic and socio-economic information that is quantified and localized. Three other benefits of the method should be mentioned. First, the group of suburbs can be included in the sample base whatever the status of the district (legal or illegal, within or beyond the administrative limit of the city). Then, due to the gain of precision from the sampling plan, the survey of a lighter sample enables us to

---

[3] An instruction manual [DUR 89] describes each of the implementation steps of the method, from the capture of the satellite image to the measure of the estimates of the information collected from the sample of households selected with this technique. Based on the exogenous information which can be mobilized to optimize the sampling plan, specific application methods are adapted.

[4] A stratified survey is a survey where the selection of the sample is done separately and independently in several sub-entities called “strata” that have been separated beforehand. If the strata have significant features with respect to the survey and are relatively homogenous, stratification will enable a precision gain on a sample of the same size [DUR 89].

observe more in depth the behaviors of the sampled households, for the same global cost. And finally, the selection of the sample on satellite image emphasizes the differences inherent to the city and favors an in depth analysis of certain sub-populations with specific behaviors. In this way, beyond satisfying the needs of localized and updated information, this method can efficiently participate in the production of knowledge on the forms and mechanisms of urbanization. A model of observed and identified reality, the satellite image is now entirely part of the information production cycle.

8.2.3. *Application in Bogota in 1993*

Bogota had over 6 million residents in 1993 when the IRD-CEDE survey on spatial mobility[5] took place. The objective of the research was to produce a two-level knowledge base on the different forms of spatial mobility of the population of Bogota and the role of these behaviors in certain transformations taking place in the Colombian capital. These two levels were: globally for the metropolitan area (forms of development) and at an intra-urban level (redistribution of densities, different forms of spatial segmentation). Due to research problems, available budget and statistical imperatives in obtaining an acceptable representation of the collected data on mobility, we have settled on the investigation of a sub-section of the metropolitan area, composed of seven areas included in the district and of four towns of the metropolitan periphery [DUR 99a]. The selection criteria for these areas aimed at showing the diversity of the metropolitan area's locations, of the socio-economic strata, of the current demographic dynamics and of the form of housing production. We now had an observation field completely adapted to a study on relations between the level of spatial mobilities as well as on urban policies and the housing market. In addition, with the two periods of the investigation (October 1993 and 1994) we have been able to analyze in a more precise way the demographic dynamics of the area in the interval separating the two observation periods. This has put us in the favorable position of analyzing in what measure the characteristics of the city and its suburbs (supply of housing, as well as economic activities) can guide or encourage certain residential practices for the individuals and their household, and, conversely, in what measure the residential practices have an impact on the suburbs' dynamics and the internal structuring of the city.

---

[5] This investigation was conducted in the context of a research program that took place from 1992 to 1996: "the forms of mobility of Bogota populations and their impact on the dynamics of the metropolitan area". Co-directed by Dureau and Florez, this program was part of a scientific cooperation agreement between IRD and CEDE (Centro de Estudios sobre Desarrollo Económico, de la Universidad de los Andes).

In each of the 11 zones that are part of the area of investigation, the sample of households was selected by the application of an area sampling plan at three levels (blocks, dwellings, households) from a SPOT image. All the images registered by SPOT Image at our request were affected by the presence of clouds. We had to use two registered scenes with an interval of four months[6] in between them, which enabled us to globally observe the study zones. It is these two images that have been the subject of processes to define the urban zone and to conduct a typology based on the density of the built-up area. The application terms of the sampling method have varied from one study zone to another. In the peripheral communities with no statistical, administrative or map information, the sample selection was only based on the satellite image. In the consolidated areas of the district, which includes better exogenous information, the degree to which the image was used varied according to the type of area, its current dynamics and the quality of the available image. For stratification in the five investigation zones within the areas and based on the density of development, we combined the classification of the image with statistical and administrative data that were available on map sections (spatial divisions grouping two dozen blocks). On the one hand, we used four indicators from the 1985 census relative to the size of the household, to the type of housing and to migration. On the other hand, we used the socio-economic stratification on six levels established by the district's administration for the application of a differential pricing of public services. It is important to note that no selection in any of the 11 zones would have been possible without using imagery.

The development of this investigation, in relation to more qualitative information collected with the other elements of the investigation system in place[7], has enabled us to highlight recent changes in the residential logic of Bogota and its metropolitan periphery (mentioned in Box 8.1), as well as to evaluate the importance of the factors determining the highly contrasted residential behaviors of the different social groups living in Bogota (access to property, family networks, distance in relation to the workplace for higher income families). The interactions between the residential practices of the people and the political actions, as well as their consequences on the development and organization of the metropolitan territory [DUR 99c] could be analyzed with the help of the information collected.

---

[6] SPOT XS images used: 12/01/93 646-340/5 L 19.6 1B and 31/10/92 646-340/5 L 26.5 1B.

[7] Besides the biographical investigation done in 1993 over 1,031 households, the investigation system includes a second passage in 1994 over the same sample, an anthropological investigation on a sub-sample of households and the collection of information on the context of each of the 11 study zones and the urban politics.

8.3. The use of imagery in urban modeling

The satellite image has only recently been used in human science modeling studies. Many applications involve urban evolution studies over long periods of time with the simulation of main development factors [BUC 99, CLA 96]; others integrate the satellite information in probabilistic diffusion models [MEA 90]. The image is viewed as a spatial location support for the analyzed phenomena (generally land occupation zones) and it provides an initial state that we can advance and compare with other images taken at different times. In another area (already mentioned in section 8.2.1), the satellite information has been used to redistribute socio-economic data based on different map units than the usual ones [MAR 91, WEB 90, WEB 92]. The use of the image in these spatial interaction models dates back to the same general period. The model presented in the next sections is an offshoot of the traditional potential model which makes it possible to integrate hypotheses on the relations which can exist between land occupation in one area and the one neighboring the same area. The use of this model can be considered from two different perspectives: static, by transferring the model on the surfaces based on the occupation of the land [DON 95] or dynamic, by associating the concept of probable "change"[8] to the concept of interaction [WEB 98]. A model of this type was applied to Bogota's urban area. By comparing the results of the model and the observed reality a few years later, we can either confirm or reject the initial hypotheses.

8.3.1. *The potential model and satellite data*

The spatial interaction models (see Chapter 1) offer the possibility of measuring the importance of the relations that exist between localized entities (markets, cities, points of sale, etc.) according to their mass (described, for example, as the population or the number of activities) and the distance separating them. The potential model uses a similar logic. It enables us to evaluate the *proximity* of each spatial entity involved in relation to a set of points, each having an influence that is proportional to its mass and inversely proportional to the distance separating it from the entity. The potential value of a spatial entity is measured from the sum of the relations between the masses of the group of points with which an interaction is possible and the respective distances to the entity. It constitutes the aggregated value of the influence of all the points of the group involved over a particular point [ABL 72].

---

[8] This concept is used in the context where the goal is to consider the difference of probable state in urban zones, including the fact that these land occupation modifications mean operational modifications of these zones, from rural to urban zones, for example.

Traditionally, this type of model was used to process flow data, such as migratory flows, for example [STE 58]. The potential of a given point measures its capacity to attract localized individuals to the other areas. Haggett [HAG 73] demonstrates the usefulness of the method in the analysis and determination of infrastructure locations such as airports or hospitals. The novelty of the works by Nadasdi and Binard [NAD 91, NAD 94] is based on a reinterpretation of this model in the context of satellite image use. The pixels (identical elements of an image or grid) are considered as the individual locations. In this way, instead of measuring the potential from a random point pattern, all the pixels of an image are considered in the calculation. The interactions between the cells represent the possible influences between land occupations in cells, according to their neighborhood.

Each pixel corresponds to a cell and its value comes from the processing of the image. Its characteristics of georeferencing, of spatial resolution (and therefore the distance between each cell center) and of affiliation to a land occupation category are known. The transformation of data used, which is radiometric at the start and then according to the various land occupation terms, enables us to associate various types of data to cells (quantitative or qualitative). In fact, initially, only the surface associated with the pixel's resolution seemed directly quantifiable. However, image processes could provide significant values after their transformation to binary information (presence or absence of built-up area), to density information (development density) according to a reference surface to define (most often the pixel), or even to ratio information that will vary depending on the numerator involved as with a vegetation index or a representation of socio-economic data [LAN 90, WEB 92].

The combination of this approach with concepts linked to image processing and color will associate several numeric surfaces resulting in the potential function. Considering that a potential surface is the result of reciprocal influences of constituting elements of the analyzed space (anthropogenic, natural, legislative or other elements), each potential surface can then illustrate, in more detail, one or the other of these aspects. Since space is considered as the product of relations between different types of elements, in order to approach the complexity of these relations we need to combine the corresponding potential surfaces. This type of combination can also provide a representation of relations between potential surfaces. These most often emphasize conflicting interaction zones between space elements: tension zones that are caused by the progression of the urbanization front, by the resistance of production modes and of space development, by the application of legislative constraints or by the existence of natural risks. The representation of a combined potential surface can then appear as an element of qualification of "possible changing areas". This combination is addressed with the theory of colors (colorimetric space[9]) and the possibilities of numeric processes. If we combine three potential surfaces, we obtain a colorful composition associating red, green and blue to each surface. The result is a colorful combination in each cell depending on the contribution of each primary color.

The results of a land occupation classification (by multivaried statistical processing of the SPOT image of Bogota taken in 1994[10]) can be used in a potential model as initial state, in order to obtain the location of spaces that may be used by urban development (residential zones, activity zones or infrastructures). They are then examined and compared with a more recent image[11]. With this comparison, we can analyze if the development potentials highlighted from the land occupation state registered are confirmed or not by an expansion or urban densification. Conclusions on residential localizations processes can be made. The gaps between the actual implementation and the modeled spaces can be from local characteristics of urban development processes and, consequently, shed new light with respect to formulated hypotheses.

8.3.2. *Application of the model to satellite imagery*

The potential model that was used answers some questions concerning the effects of compound influence of a distance and a quality, land occupation. Each land occupation will be given a weight that will play the part of the traditional "mass" in potential models.

Land occupation is a matter of defining which cells are "close" to one another from their land occupation. This proximity reflects spatial as well as qualitative (the affiliation to one group) proximity. The hypothesis is that the measure of this proximity represents the intensity of the interactions that exist between two locations characterized by similar or different land occupation categories. The necessary quantification of the data is done by the determination of a series of weights (weight occurring at numerator) depending on the theme of the study and applied to categories of land occupation. They are ranges of values that show land influence (built up surface and reference surface) or the intensity of land occupation (relation between floor surface and a reference surface allocated to buildings) [WEB 98] in land development. In the vegetation sector, the association with biomass is often used, but the consideration of the production is also possible.

---

[9] Space in which color is considered as a vector whose trichromatic coordinates are read in axes that define the type of space: red, green and blue (RGB space) or intensity, tint and saturation (ITS). Any vector combination of color then becomes possible within one space.

[10] References of SPOT XS image from 1994: 24/06/94 647-340/7 L 27.2 1B.

[11] References of SPOT XS image from 1998: 02/01/98 647-340/4 R 27.1 1B.

The general form of the model consists of measuring the potential including an auto-potential effect for each of the n cells. In this way, the potential in a point i corresponds to the sum of the effects of n locations over i and the effects of i over itself:

$$\mathrm{Pot}^{\alpha}_{i} = a\,\mathrm{M}^{\alpha}_{i}/(1/2\ \mathrm{S}_{i}^{\ 1/2}/\pi) + a\,\mathrm{M}^{\alpha}_{1}/\mathrm{d}_{i1}^{\ b} + \ldots + a\,\mathrm{M}^{\alpha}_{i-1}/d_{\,i,i-1}^{\ b} + \ldots + a\,\mathrm{M}^{\alpha}_{i+1}/\mathrm{d}_{i,i+1}^{\ b}$$

with n = number of pixels of the image, $j = 1, n$ and $j \neq i$.

In other words:

$$\mathrm{Pot}^{\alpha}_{i} = a\,(\mathrm{M}^{\alpha}_{i}/(1/2\ \mathrm{S}_{i}^{\ 1/2}/\pi) + \textstyle\sum_{j} \mathrm{M}^{\alpha}_{j}/\mathrm{d}_{ij}^{\ b})$$

with:

$\mathrm{Pot}^{\alpha}_{i}$ = potential of i for the occupation of land α;

S_{i} = surface of pixel i in the Bogota example: 20*20m;

M^{α}_{i} = weight applied in i for the occupation of land α;

b = exponent explaining the scope of the decrease of the influence of a cell over another when the distance separating them increases;

a = constant (often unit);

d_{ij} = distance between two cells (pixels) i and j;

α = category of occupation of land (classes 1, …, n).

The introduction of the exponent α characterizes the transfer from the "traditional" model to satellite imagery. A value of α corresponds to each land occupation class. Thus, applied to all the pixels of a satellite image, the potential model will enable us to locate and quantify the zones where land occupation zones are probable (dense zones becoming even more dense, zones of dispersed habitat filled up by spreading of proximity) from the hypothesis put forward (densification and expansion of the developed area, for example). In this way, the potential model is calculated for all the pixels while taking into account all the categories of land occupation.

In order to measure the potential function over the whole image, it is necessary to make a transfer from the result of an image processing (in this case of a land occupation classification) to a statistical surface resulting from the measure of potential. This is equivalent to applying a "filter" on all the pixels of the image by using a sliding window[12] in which the calculation is done by incrementing each pixel, and the result affects the central pixel of the window. The surface that is then obtained has topological properties and can be interpreted in terms of gradient and flow [DON 95]. To minimize the border effects, we have preferred the application of a circular sliding window.

In the application of the function of potential, the window size, the exponent choice and the weights are particularly important [ROU 98, WEB 98]. The size of the window is the result of a compromise between the objectives in terms of detail, of variety of land occupation types and of the cost of the calculation. It must be large enough to consider the biggest possible variety of land occupation categories, but small enough to cover the central and peripheral structures. In addition, there is a direct relation between the size of the window and the results: the larger the window, the more pixels are taken into account for the calculation. The exponent of the distance acts as a smoothing element that is more or less effective on the values: if the value is higher than 1, then the neighboring pixels will have a higher influence, otherwise the smoothing effect is more significant. A tricky decision resides in the characterization of the weights based on land occupation categories. The qualitative value of the pixel is weighted according to the objectives of the study, of field knowledge and the range of values used. If we consider in fact that the result corresponds more specifically to the illustration of the possible effects of urbanization processes, i.e. to the interaction between "urban" surfaces, then the weight values will be higher for the categories that qualify urban development than for natural or agricultural zones. If, on the contrary, the hypotheses are based on a resistance from natural zones – they may be protected, for example – the corresponding categories will have higher values. The information concerning land area or the intensity of land occupation, the types of surface developing, and even the relative importance of the concerned surfaces are often used to weigh the different categories of land occupation. These weights must still be completely consistent with the objectives, as with the initial configurations of the spatial elements. The proposed methodology, which consists of considering several potential surfaces, is equivalent to proceeding to an integration of information on the basis of a combination. The association of several representations combining the hypotheses of possible development of urban space is entirely possible in the context of GIS or of image processing. The representations can be associated with intensities of interaction co-occurrence or be the subject of a more sophisticated definition that may come from a multicriteria analysis, for example.

---

[12] A spatial filter that applies on all the pixels of an image a function that is calculated over all the cells in a window of variable size and form. The result affects the central pixel of the window. This window is called "sliding" because it moves from pixel to pixel along the lines of the image.

8.3.3. *Application in Bogota*

In order to verify the validity of the model, we compare the results in terms of possible expansion zones and actual land modifications. Two images of Bogota have been used: one from 1994 gives the initial situation of the urban area and enables us to make hypotheses on the evolution dynamics which could transform the terms of production of space. The other, which is more recent (1998), offers the possibility of understanding the evolutions that have taken place between both reference times. This provides an appreciation for the zones in mutation and the intensity of localized movements. The processes can be spatially illustrated. In addition, compared with the zones highlighted by the potential model of 1994, the results of the image classification of 1998 helps us verify the hypotheses made with the choice of weights and to identify the gaps, perceptible and which can be located, in relation to the outline of development that seemed relevant.

The classifications obtained from multivariate statistical processes inform us on the major features of land occupation in the northern part of Bogota. We can localize the stable masses of the city and the forms resulting from the development four years later. These structures, emphasized with the help of categories of occupation and land use, are the product of multiple and combined processes operating on various temporalities. The space production and usage processes are highlighted through spatial mutations from one state to another (for example, the construction of an airport) or through transitions from one mode of agricultural development to another. The importance of the mutations is an indicator of the intensity of the urban space production process that the new developed zones illustrate.

The images then show a spatial organization at two moments. The image from 1994 (see Figure 8.2) emphasizes, on the one hand, a dense core in the southeast that is delimited by mountainous foothills to the east and a small central spur and, on the other hand, a large agricultural plain to the northwest, with one specific element: some horticultural greenhouses, sometimes large, generally close to urban fringes. This zone corresponds to a habitat intended for higher income families, whereas the western part of the capital is traditionally occupied by middle-income families. Infrastructures are clearly distinguishable: an airport to the west, large north-south axes of circulation, as well as parks and leisure activity centers within the urban fabric, in particular toward the center and the east.

Given the present taxation pressure in a city with a high growth rate, we can wonder about the role of greenhouses in the urban expansion process of the Colombian capital. We can in this way formulate the hypothesis of an “attractive” role of the greenhouses: these can constitute an interesting indicator of the dynamics of territorial reconstruction and become issues in the modification of modes of space development from the moment that the added value of the horticultural production cannot compete with taxation pressure anymore. The potential model has then been weighted according to these prerequisites and the results of the model compared with the actual situation in 1998.

From the comparison of the results of the two classifications, we can analyze land occupation changes and obtain a first spatial analysis of the urbanization and densification processes (see Figure 8.2).

In light of the data obtained, the changing rates tend to show that there has been growth in general in mineral built-up surfaces (construction that is very dense, dense, or weak with vegetation) and that the agricultural surfaces have actually decreased and that the greenhouses are also decreasing by 13% (see Table 8.2).

However, several nuances must be taken into account with regard to these results. The modifications can correspond to urban reality: this is the case with the densification of the center and the development of urban fringes west of the zone. However, they can also be attributed to agricultural practices modifying the agricultural landscape following a production schema, or to phenologic differences due to the dates of image capture. Finally, they can come from classification errors due to high radiometric proximity of certain elements: bare land is spectrally close to a building; the values obtained with the different sensors will be similar.

The application of the potential model is interesting for two reasons:

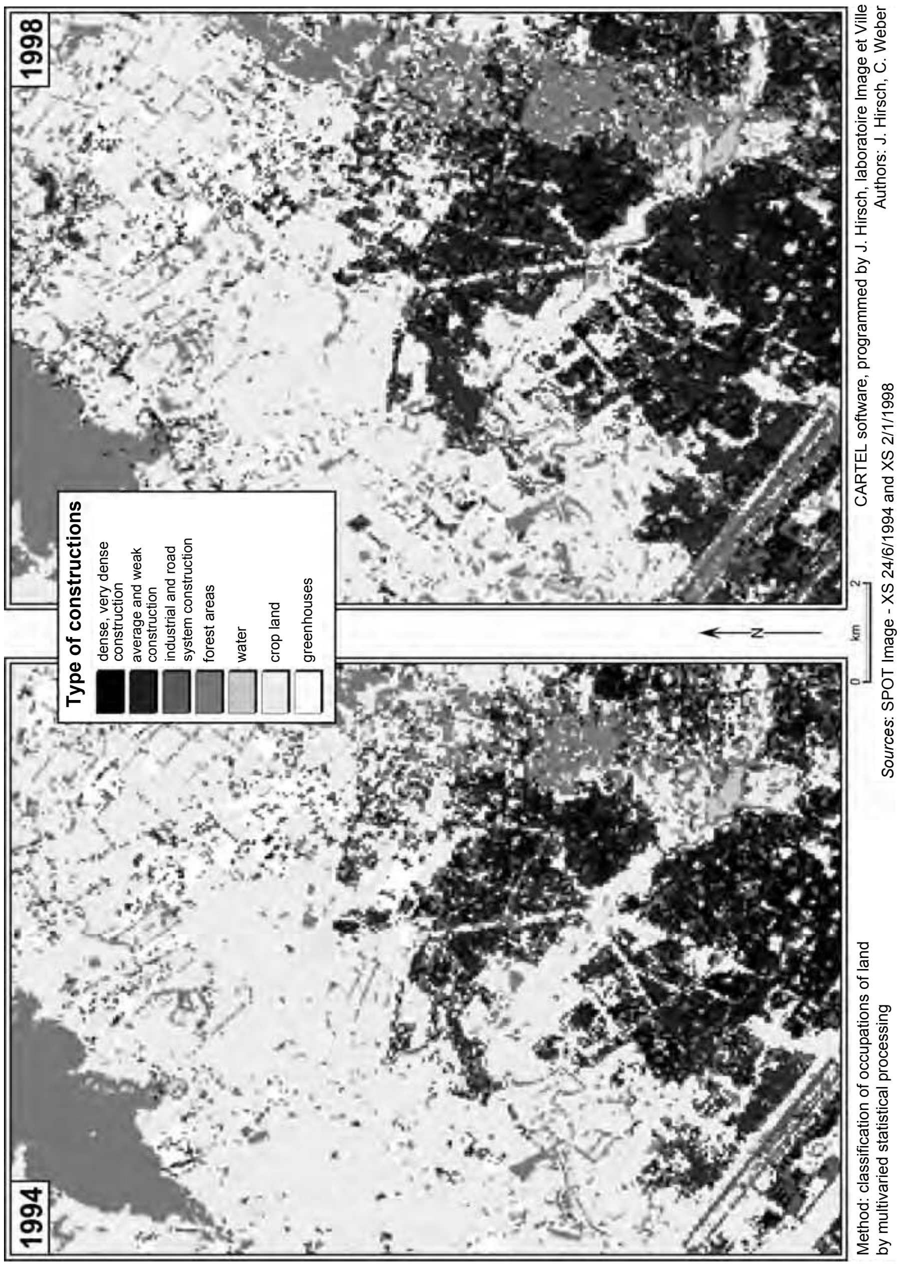
– the possibility of localizing sensitive zones, which are subjected to change due to their intermediate position between two types of land use/cover. Thus, a zone located between dense constructions and a residential zone will have a high probability of densification;

– it contributes to the confirmation or rejection, from a visualization of possible changes, of hypotheses expressed *a priori* and to thinking about the consequences on spatial distributions.

The chosen weights for the three calculated potential surfaces are presented in Table 8.3. The obtained representations in the case of Bogota correspond to the weight choices – defined by the density of the built-up areas and to their use/cover densities – to types of vegetation surfaces and to modes of agricultural production and, finally, to the directions associated with the development of the city (administrative or natural constraints, activities hypotheses, etc.). The values attributed to weights come from the internal logic of each potential surface. It corresponds to the land use/cover rate of the category involved.

Classification name	**Order**	**Surface in 1994**	**Surface in 1998**	**1998/94 difference**	**Relative variation 1998/94**
Very dense construction	1	74,021	89,894	15,873	+ 21.4%
Dense construction	2	26,205	24,374	–1,831	– 7%
Weak construction	3	120,799	160,205	39,406	+ 33%
Industrial and road system construction	4	30,976	34,990	4,014	+ 13%
Almost bare and crop land	5	653,795	530,417	– 123,378	– 19%
Water	6	1,168	1,747	579	+ 49.5%
Greenhouses	7	60,371	52,365	– 8,006	**– 13%**
Forest areas	8	243,525	316,868	73,343	+ 30%

Table 8.2. *Categories of occupation of land from the classification of SPOT images 1994 and 1998*



Method: classification of occupations of land by multivaried statistical processing

Sources: SPOT Image - XS 24/6/1994 and XS 2/1/1998

CARTEL software, programmed by J. Hirsch, laboratoire Image et Ville

Authors: J. Hirsch, C. Weber

Figure 8.2. *Bogota (Colombia): evolution of urban area between 1994 and 1998 (SPOT images classification)*

For the “urban potential” surface, the choice of high values for dense and very dense built-up classes is analyzed according to the coverage of land occupations in an urban environment. For the “weakly built-up” class which corresponds to a residential land occupation with green spaces, the affected value characterizes a more loosely spaced structure.

In the same way, the weights associated with agricultural or forest spaces correspond to a rate of agricultural production that is separated from the phenologic state discernable on the satellite image.

The third surface corresponds to the hypotheses made about the growth dynamics of Bogota: a suburban increase constrained by the topographical characteristics and a dispersion by contiguity for certain urban production zones such as along the communication axes. The hypothesis of an attractive role of the greenhouses results in the attribution of a high value (7). Residential development is tricky in this part of Bogota: with wealthy families it will lead to stronger competition for pieces of land that are closer to the city.

Classification name	**Order**	**Urban surface**	**Vegetation surface**	**Development surface**
Very dense construction	1	**10**	0	0
Dense construction	2	**9**	1	1
Weak construction	3	**8**	**3**	**4**
Industrial and road system construction	4	**9**	1	1
Almost bare and crop land	5	0	**8**	**5**
Water	6	0	0	0
Greenhouses	7	0	**4**	**7**
Forest areas	8	0	1	1

Table 8.3. *Weights used for the potential model*

The surfaces obtained are then associated according to a colored composition which makes it possible to isolate the actual or competing influence zones from the different potential surfaces. This result is obtained by classifying the values obtained for the three surfaces according to a threshold that makes it possible to isolate the behaviors of one zone or another based on the identifiable characteristics of the urban, vegetation or development potential. Table 8.4 summarizes this classification.

Urban potential	**Vegetation potential**	**Development potential**
Strong construction	Average	Low
Average construction	Average	Average
Weak construction	Average	Strong
Weak construction	Strong	Strong
Weak construction	Average	Average

Table 8.4. *Classification of potential surfaces*

With the comparison of the results and the image from 1998, we have been able to verify the hypotheses formulated for probable development zones and the role played by greenhouses. Though the urban developments have occurred as assumed in the zones that were highlighted as potential surfaces, the hypothesis of the attractive role of greenhouses must be rejected. In fact, even if constructions were located in some areas, the general resistance of greenhouses to urban development is spatially verifiable in high-income areas as well as middle-class areas.

There are several considerations associated with this comparison:

– they obviously involve the solidity of the initial information, i.e. the classifications of 1994 and 1998: artifacts have been introduced due to bare agricultural land or greenhouses whose aerial structures, most often made of plastic in Bogota, have been taken out or put back between both dates;

– we can question the identified development dynamics factors: Figure 8.3 clearly shows that the densification of central districts comes with a sustained suburban expansion process. Similarly, in the northern region of Chia, the expansion largely infringes on agricultural zones and the increase of road infrastructures suggests future development along the northeast-southwest highway. Certain anomalies are visible, such as the airport, which was difficult to predict from the satellite data of 1994;

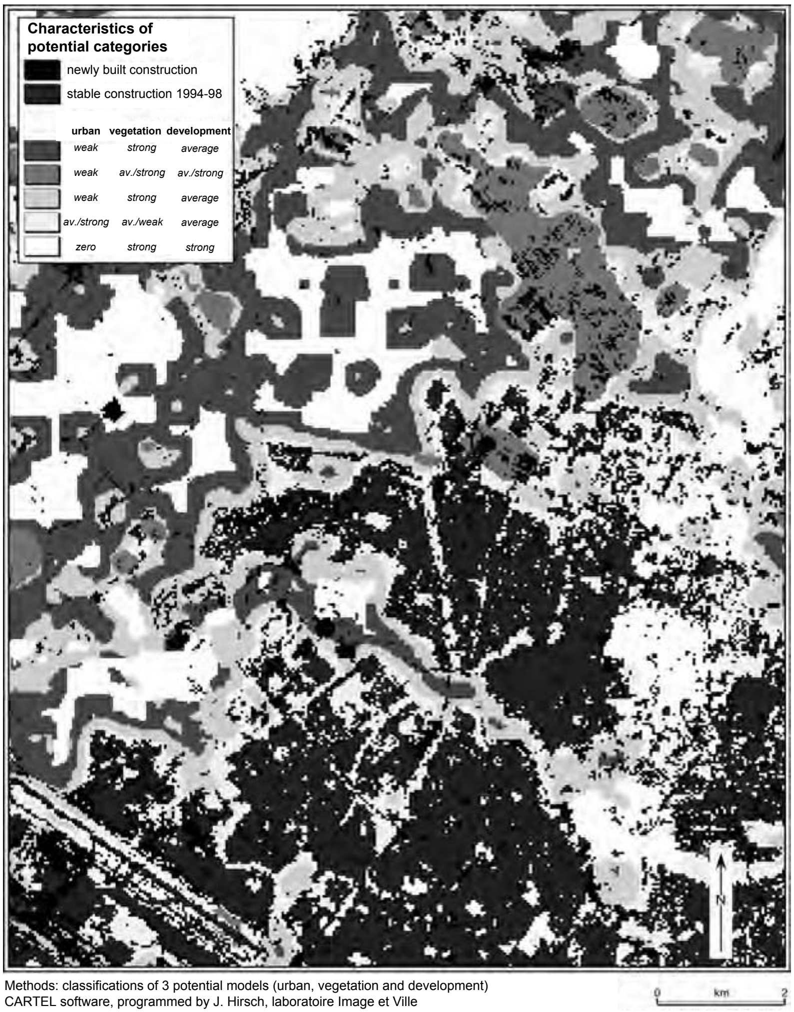
– finally, the introduction of imagery and its demographic results (by area sampling) and localizations of potential development will be analyzed within the integration into a dynamic model linking residential mobility forms and territorial reconstruction.

The spatial comparison possibilities enable us to look beyond the digital results of a statistical comparison and visualize, for example, zones subjected to competition. In reading Box 8.1, we could base a hypothesis that would have been very difficult to evaluate without spatial information (image, aerial or land photograph). Here again, the image contributes to spatial modeling by positioning the phenomena and by encouraging other considerations on space and its possible changes.

8.4. Spatial information and dynamic modeling

The preceding pages have shown that the major cities, whether they are located in industrialized or developing countries, require adapted and reliable informational devices enabling the capture of their growing complexity and of the dynamics in place. These are complex, open systems, far from equilibrium, for which the auto-organization paradigm seems to have taken over. However, this approach requires that spatio-temporal relations between observation levels (individual at microscopic level and organization at macroscopic level) are emphasized. The connections between single event and global structure remain a major area that still needs to be studied: they associate with a particular geographical object that is not the individual or the group but a *geographical configuration*, i.e. an organized and localized dynamic entity, with its own life cycle. Considering space as the product of relations between elements instead of a simple support, leads to the modification of the usual approaches and to reflect on the intrinsic nature of these entities: space and time [PUM 89]. Traceable and identifiable, the entities must also be tracked (trajectory) and evaluated according to their degree of adaptability or resistance to change in such a way that the links between observation levels are maintained. This approach requires a constant positioning in relation to the objectives that the constituted informational device must reach and their relevance.

In the example presented here, the informational device is particular. On the one hand, the demographic data give us information on individuals, their family and professional biography and their residential trajectory, and helps us derive rules of residential location choice. On the other hand, the results from the potential model can be reused to characterize the urban zones providing information on the possibilities (or non-possibilities) of urban space development (densification, expansion). The goal is to draw a dynamic model (SIMBOGOTA) which integrates the operation of the potential model previously described. This model is currently being developed[13] and we will only give an overview here.



Methods: classifications of 3 potential models (urban, vegetation and development)
CARTEL software, programmed by J. Hirsch, laboratoire Image et Ville
Sources: SPOT XS 1994 and XS 1998 image

Authors: J. Hirsch, C. Weber

Figure 8.3. *Bogota (Colombia): comparison of potential models 1994 and of construction 1998*

---

[13] Studies done by a multi-institutional team (CNRS, IRD and Universities), with financial support from PSIG-1998 (see [WEB 99]) and from ACIV-1999 (see [DUR 00]). A dissertation from DEA [VAN 99] was developed in this context.

8.4.1. *Towards a dynamic multilevel model*

The connection between a socio-demographic approach and a spatial approach of the urban system seems to be one way to respond to the many questions on dynamic processes within cities. The territorial organization of a major city, at the center as in its periphery, corresponds to processes whose effects are spatially traceable. Spatial units identified locally can be associated with the movements of the population studied (by survey in the case of Bogota). Globally, the effects of these movements are often generalized within these zones which could evolve from a land use/cover state to another state according to the evolution processes in place (for example, densification of built-up areas or new constructions). These mutations can then be assessed from a dynamics point of view (change analysis). It will then be important to develop a model that will connect the individual mobilities, changes of housing stock, modifications of land use, as well as the interactions between the different associated spatio-temporal levels.

The hypothesis is that satellite data enables us to make the connection between the urban evolution dynamic processes occurring at a global spatial organization level (territorial expansion and reconstruction) and the individual mobility trajectories occurring at local level. The goal is that the localization and identification of development zones can be addressed from the comparison of two states (two satellite images or results from processing and additional information) and explain the processes associated with urban expansion and internal territorial reorganization in the capital based on an adapted taxonomy. The dynamic model will then have to combine the results from the potential model and the rules operating at individual level.

8.4.2. *Application in Bogota: a preliminary simulation*

Bogota is considered as the product of the behavior of its population, so the objective is to analyze the way in which the different spatial mobility forms influence the organization and the use of the space in the city and are modified in return. The mobility is therefore designed as a factor that is determined and determining at the same time by the dynamics and the territorial reorganization observed through the urban discontinuities, segregative processes (within their different dimensions: socio-economic, demographic, ethnic, etc.) and the peripheral expansion and density redistribution processes within the metropolitan space. The connections between mobility behaviors (residential and daily) of individuals and households and the emergence of new organization forms of the metropolitan territory are at the core of our questioning. In order to understand and respond, we favor an approach in terms of dynamic modeling in addition to statistical approaches. A certain number of hypotheses have been raised in the context of studies already available. The MASs' *in fine* choice (see Chapter 7) for modeling is explained by the necessity of integrating space as well as human elements in the model.

In order to develop such a model, questioning must be done on several points:

– the residential system (location, observation levels) and its concepts (proximity, polarization, cognition);

– the decision-making system (individual, group, household or decision-making unit) linked to behavior rules obtained by land observations (investigations conducted): forms of mobility have been identified and characterized according to a driving process, for example, the residential aptitude (capability to choose);

– the developed driving force of mobility is based on the evaluation of well-being, the satisfaction with the residence area. The determining factors of mobility (social status, comfort, proximity to workplace, neighborhood or network relations) have been introduced by two concepts, "homophilia or homophobia", in order to explain segregation mechanisms brought about by the research or the avoidance of individuals from certain groups. A satisfaction factor, measured for each individual, will serve to evaluate his propensity to migrate.

There are still many more points to develop such as, for example, the receiving capacity for migrants in spatial units. The definition of these capacities can depend on the maximum number of residences, the rate of construction, the construction density or the aptitude to prefer residential implementations (from the identification of probable development zones resulting from the potential model).

The information obtained on possible changing surfaces can then be used in other ways, in particular, to qualify "future" spaces within dynamic modeling where it is necessary to establish probable spatial development outlines. The potential model gives a necessary indication for the initialization of a process of movement, the probable change zones, knowing that this information is not restrained by an administrative units set and that it can therefore consider "spaces out of urban administrative delineations". There are still many points to address like, for example, the structure of information and their spatio-temporal follow-up. The passage of a potential surface to an identifiable spatial entity by an "agent" as a possible migration area will certainly consider the proximity of the locations, the qualification of the sites, as well as the inaccuracy of their spatial limits and therefore the configuration of the agents "space or spatial unit" based on the choices retained until now.

8.5. Conclusion

This chapter will have highlighted, with the help of three examples, the diversity of roles that satellite images can play. This source of information can occur efficiently at different levels in a process of knowledge production and modeling. The example developed throughout the chapter concerned the interpretation of a particularly complex system: flow system and place system at the same time, the city constitutes an ideal case for the demonstration of the possibilities of satellite imagery in social sciences. Basically digital information, satellite imagery lends itself particularly well to an integration in traditional or current modeling procedures, including those that consider spatio-temporal processes with interaction.

In the beginning of spatial remote sensing, there were several applications that would isolate this source of information from the existing informational context. This approach was no doubt caused by the relatively exploratory character of its first uses and by a will combined with caution to evaluate what this new source of information would bring in the long term. Through time and due to its increasing know-how and to the development of geographical information systems, it has been made rather clear that satellite imagery increases its value when it is associated with other information and used at different stages of production and analysis of the information. The demographic applications presented here provide a convincing example, although current image resolutions (SPOT) still constitute an important limit. With its resolution of 20 meters in multispectral mode and of 10 meters in panchromatic mode, the SPOT satellite, launched in the mid-1980s, has marked the beginning of a new era in urban applications of spatial remote sensing.

A new phase is emerging with very high spatial resolution images. With a resolution of 4 meters in multispectral mode and 1 meter in panchromatic mode, the Ikonos satellite provides documents very close to traditional aerial photography. These new capabilities open wide application horizons. Expectations from users such as territorial communities, space management and environmental protection organizations, road planning and telecommunications companies, etc. are equally wide. The responses to needs for precision, updating and cost optimization will be analyzed in great detail. We can hazard a guess that the already significant diversity of remote sensing uses in research and modeling approaches, partly illustrated in this chapter, will only increase in the future.

8.6. Bibliography

[ABL 72] ABLER R., ADAM J.S., GOULD P., *Spatial Organisation. The Geographer's view of the World*, 1972.

[AVE 93] AVERY T.E., BERLIN G.L., *Fundamentals of Remote Sensing and Airphoto Interpretation*, Macmillan, New York, 1993.

[BAT 95] BATTY M., "The computable city", *Fourth International Conference on Computers in Urban Planning and Urban Management*, Melbourne, 1995 (can be viewed at http://www.geog.buffalo.edu/Geo666/batty/merlbourne.html).

[BUC 99] BUCHANAN J., ACEVEDO W., "Defining the temporal and geographic limits for an urban mapping study", http://edcwww2.cr.usgs.gov/umap/htmls/pubs.html, 1998.

[CLA 96] CLARKE K.C., HOPPEN S., GAYDOS L.J., "Methods and techniques for rigorous calibration of a cellular automaton model of urban growth", http://geo.arc.nasan.gov/ usgs/clarke/calib.paper.html, 1996

[COU 85] COUCLELIS H., "Cellular worlds, a framework for modelling micro-macro dynamics", *Environment and Planning B*, 17, p. 585-596, 1985.

[CUR 97] CURRY M.R., "The Digital Individual and Private Realm", *Annals of the Association of American Geographers*, 87, 4, p. 681-699, 1997.

[DES 93] DESPRATS J.F., DUTARTRE P., *Estimation de la population des quartiers périphériques de Ouagadougou par télédétection*, Orleans, BRGM, 1993.

[DON 95] DONNAY J.P., "Délimitation de l'hinterland des agglomérations urbaines au départ d'une image de télédétection", *Revue Belge de Géographie*, p. 325-331, 1995.

[DUR 89] DUREAU F., BARBARY O., MICHEL A., LORTIC B., *Sondages aréolaires sur image satellite pour des enquêtes socio-démographiques en milieu urbain. Manuel de formation* (3 versions: French, English, Spanish), Paris, ORSTOM, Didactique Collection, 1989.

[DUR 95] DUREAU F., WEBER C. (eds.), *Télédétection et systèmes d'information urbains*, Anthropos, Villes Collection, Paris, 1995.

[DUR 97] DUREAU F., "La production rapide d'informations démographiques et économiques par sondage aréolaire sur image satellitaire: application à trois villes de pays en développement (Quito, Bogota et Yaoundé)", in *Télédétection des milieux urbains et périurbains,* Dubois J.M., Donnay J.P., Ozer A., Boivin F., Lavoie A. (eds.), AUPELF-UREF, p. 215-224, 1997.

[DUR 99a] DUREAU F., FLOREZ C.E., "Enquêtes de mobilité spatiale à Bogota et dans 3 villes du Casanare (Colombie)", in *Biographies d'enquête. Bilan de 14 collectes biographiques*, Collective, INED-PUF, Méthodes et savoirs Coll., Paris, p. 241-278, 1999.

[DUR 99b] DUREAU F., "La mobilité à géométrie variable des habitants de Bogota", *Espaces, Populations, Sociétés*, 1999-2, p. 329-344, 1999.

[DUR 99c] DUREAU F., LULLE T., "Le développement spatial de Bogota durant les 20 dernières années: le rôle direct et indirect des politiques urbaines", in *Les grandes villes d'Amérique Latine: quelles politiques en cette fin de millénaire ?,* Thomas F. (ed.), special number of *Revue de Géographie de Lyon, Geocarrefour,* no. 4, p. 291-300, 1999.

[DUR 00] DUREAU F., TREUIL J.P. (eds.), Mobilités et dynamiques urbaines: une application en intelligence artificielle distribuée, Project submitted to Action incitative Ville, 2000.

[FOR 85] FORESTER B.C., "An examination of some problems and solution in monitoring Urban Areas from Satellite Platforms", *International Journal of Remote Sensing*, vol. 6, p. 139-151, 1985.

[GRE 56] GREEN N.E., "Aerial photographic analysis of residential neighborhoods: an evaluation of data accuracy", *Social forces*, no. 35, p. 142-147, 1956.

[HAA 97] HAACK B., GUPTILL S., HOLZ R., JAMPOLER S., JENSEN J., WELCH R., Chapter 15 in *Urban Analysis and Planning. Manual of Photographic Interpretation*, American Society for Photogrammetry and remote Sensing, Bethesda, Maryland, p. 517-553, 1997.

[HAG 73] HAGGET P., *L'analyse spatiale en géographie humaine*, Armand Colin, 1973.

[IAU 83] IAURIF, ROC, SIDAU, SEE, *Les données urbaines. Système permanent de production rapide et économique. Recherche et synthèse*, IAURIF, Paris, 1983.

[JEN 98] JENSEN J.R., HALL J., MICHEL J., "A systems approaches to Environmental Sensitivity Index (ESI). Mapping for oil spill contingency planning and response", *Photogrammetric Engineering & Remote Sensing*, vol. 64, no. 10, p. 1003-1014, 1998.

[JEN 99] JENSEN J.R., COWEN D.C., "Remote Sensing of Urban/Suburban Infrastructure and Socio-Economic Attributes", *Photogrammetric Engineering and Remote Sensing*, vol. 65, no. 5, p. 611-622, 1999.

[LAN 90] LANGFORD M., UNWIN D.J., MAGUIRE D.J., "Generating improved population density maps in an integrated GIS", *Proceedings of EGIS'90*, vol. 1, p. 651-660, 1990.

[LIN 85] LINDGREEN D.T., *Land Use Planning and Remote Sensing*, Martinus Nijhoff Inc, Boston, 1985.

[LO 86] LO C.P., *The Human Population. Applied Remote Sensing*, Longman, New York, p. 40-70, 1986.

[MAR 91] MARTIN D., "Understanding socio-economic geography from the analysis of surface form", *Proceedings of EGIS'91*, vol. 2, p. 691-699, 1991.

[MEA 90] MEAILLE R., WALD L., "Using geographical information system and satellite imagery within a numerical simulation of regional urban growth", *International Journal of GIS*, vol. 4, no. 4, p. 445-456, 1990.

[NAD 91] NADASDI I., BINARD M., DONNAY J.P., "Transcription des usages du sol par le modèle de potentiel", *Mappemonde*, 3/91, p. 27-31, 1991.

[NAD 94] NADASDI I., BINARD M., "Toward a rational utilisation of satellite imagery for Lisbon's metropolitan planning", *EARSeL International Journal. Advance in Remote Sensing*, vol. 2, no. 3, p. 130-137, 1994.

[PHI 97] PHIPPS M., LANGLOIS A., *Automates cellulaires: applications à la simulation urbaine*, Hermès, Paris, 1997.

[POI 97] POIRIER J., PÉTILLON L., "La télédétection et la démographie ou l'initiation d'une équipe à une nouvelle méthodologie de sondage", *Communication at 23rd UIESP Conference*, Beijing, 1997.

[PUM 89] PUMAIN D., SAINT JULIEN T., SANDERS L., *Villes et auto-organisation*, Economica, Paris, 1989.

[ROU 98] ROUSSELET B., *Calibrage du modèle de potentiel aux images de télédétection. Applications urbaines et régionales*, Dissertation from Liège University, Geographic Sciences, 1998.

[STE 58] STEWART J.Q., WARNTZ W., "Physics of population distribution", *Journal of Regional Science*, vol. 1, p. 99-123, 1958.

[VAN 99] VANBERGUE D., Modélisation des migrations intra-urbaines. Approche multi-agents. Application à la ville de Bogota, Dissertation of DEA, Paris University 6, 1999.

[VER 78] VERNIÈRE M., "Méthode d'analyse quantitative de la croissance urbaine dans l'espace et dans le temps. Exemple d'une banlieue de Dakar (Sénégal)", *Photo-interprétation*, no. 1, fasc. 4-5a, Technip Editions, Paris, p. 34-55, 1978.

[WAT 85] WATKINS J.F., "Small area population estimates using aerial photography", *Photogrammetric Engineering and Remote Sensing*, vol. 51, n° 2, p. 1933-1935, 1985.

[WEB 90] WEBER C., HIRSCH J., "Classification pixel et Classification par zone: Tests sur image SPOT en milieu urbain", in *Caractérisation et suivi des milieux terrestres en régions arides et tropicales*, *2es Journées de Remote sensing ORSTOM, December 4-6, Bondy,* ORSTOM Editions, p. 231-239, 1990.

[WEB 92] WEBER C., Traitement de l'Information Satellitaire et Modélisation Urbaine: Contraintes de discrimination et reproductivité, National Spatial Remote sensing program – 90N50/0074, CNES/CNRS, 1992.

[WEB 95] WEBER C., *Images satellites et milieu urbain*, Hermès, Paris, 1995.

[WEB 98] WEBER C., "La croissance urbaine de Kavala, évolution et perspectives", *Bulletin SFPT*, no. 151 (1998-3), p. 29-37, 1998.

[WEB 99] WEBER C., Modélisation dynamique des interactions entre Forme de mobilité et recomposition territoriale, PSIG report, 1999.

Chapter 9

Mathematical Formalization for Spatial Interactions

9.1. Introduction

Several phenomena are modeled by what is commonly called spatio-temporal dynamics. These are processes for which information localization in space plays a specific role. Each of the disciplines like economy, sociology, ecology, etc. has the opportunity to consider or not this type of information. For example, since the end of the 19th century and with the works of Von Thünen, we have seen the emergence of spatial economy, where the localization of socio-economic players is part of the models. Many other examples will be given throughout this chapter, but we can already envision networks of cities, forests considered as tree networks, or even a eukaryotic cell whose operation is based on a compartmentalization of biochemical processes. We could imagine that space is simply one more variable that we would need to consider, but “things” are a little bit more complicated: space plays a particular role in modeling.

Modeling a natural phenomenon often comes down to formalizing, in a specific and simplified language, the mathematics of what we usually describe in discursive and more complex form. The art of modeling is finding the best compromise between the simplicity of the mathematical form and the respect of prominent and identity elements of the phenomenon studied. This compromise wins the joint support of mathematicians and specialists of a natural discipline with respect to the relevance of a mathematical language based approach. However, there is a problem

---

Chapter written by Alain FRANC.

with reaching a compromise from a mathematical language: we must use a language that is based on accuracy and that will tend to eliminate any approximation, imprecision or ambiguity in reaching a *compromise* with other disciplines in the study of a natural object, as each discipline has its own viewpoint [BOU 99]. The combination of mathematical accuracy and negotiation to finally reach an accepted compromise, after a long exploratory series of trials and errors, is not instantaneous in terms of results or Cartesian in terms of method… If developing and using a rigorous language constitutes the contribution of mathematics, building the compromise cannot be done simply with that discipline's rigorous formalizations. Mathematics is not a science of compromise.

The first step in mathematical modeling often consists of quantifying the description of the objects studied. It is important to note here that this step is not exclusive. The mathematical language enables us to process qualitative descriptions as well, such as the use of algebraic structures in family ties [RIC 71]. This quantification associates a list of variables to each object studied which, as the first step in the modeling, simplifies its description without distorting it. The simplest examples are the area of a territory, which considers its size but ignores the spatial organization of the land, or, more naturally, every global indicator of human activity like GNI (gross national income), etc. Often presented as "objective", quantification is actually implicitly linked to the type of compromise chosen. However, there is not an only way of quantifying an object, or even a better compromise for a question raised about an object. The compromise naturally depends on the discipline used, which will probably guide us toward a type of unit of measure, or of quantification. In ecology, for example, a dendrometry specialist that is interested in growth will measure the size of a tree by its height in meters, whereas a physiologist will measure its biomass in grams of carbon stored in the tissues. This quantification is a question of compromise between object and question, but it is also a question of habit: the choice of relevant variables is often the cornerstone of a discipline and is a quantitative description choice which brings together specialists of one discipline.

We can imagine an object, for example a network of cities, represented by the population of each city within the network. Specifically, the description of this network is a vector whose k-th component is the population of city k. We imagine, but it is only an example, that the question studied is the evolution of these populations. We know the population numbers in 2000 and we are attempting to find an evolution model that will predict these populations in 2020. This evolution will take the endogenous dynamics of each of these cities into account, as well as the exchanges between them. The mathematical formalization that naturally comes to mind is that of a dynamical system: we know state $u(t)$ of a system at moment t (this state can be the vector containing the population of the cities), we have a measurement rule to calculate state $u(t + 1)$ at moment $t + 1$. We can gradually calculate the state at any moment $t' > 0$.

The question that instantly comes to mind is: does the position of the cities in space influence this evolution rule? In other terms, in order to calculate state $u(t + 1)$ from $u(t)$ is it necessary to know the spatial links between cities? The answer is in general affirmative for the unit systems in interaction, as in cities in a network: space is an element of the dynamics. Its status is, however, tricky. If, from a mathematical viewpoint, we give each city at position (x,y) the status of a variable, by simply deciding that each city is represented by a triple (u_k,x_k,y_k), we also perceive that the status of space is not rigorously identical to the status of a variable of state. In fact, space represents a context within which the dynamics takes place: the evolution of the system is a play and space is its theatre. We can propose a second example of this duality from argument modes in silviculture. A tree grows from resources exchanged with the environment: light (photons trigger the photosynthesis reaction), atmospheric carbon, water and mineral elements. Physically, we can follow the exchanges of these matters as we follow the exchanges of populations between cities. However, it is easy to convince ourselves that the quantity of resources that a tree can use is proportional to the space that it uses for its architectural development: the bigger the roots, the more developed the leaves, and the more significant the quantities of light, carbon dioxide, water and nutrients absorbed. These quantities (foliar surface, combined roots lengths) are on a scale relation with a surface that is the projection on the ground of the tree crown. In a growth equation, we can then mathematically replace the quantity of resources by space available for growth [FOR 92]. However, the space is not a variable of state measuring a quantity that circulates and exchanges: it is an indicator of growth possibilities. In other words, it is not an actor but the action theater. It would be excessive to consider space as a resource, actor of growth, even if the role that space plays in the equations is that of a resource.

Even though we can define geography, starting with its etymology, as a science of space, the consideration of space in dynamics is not specific to geography and, in this section, we will give some examples of the relevance of this problem beyond this field of discipline, in domains that, when regularly connected with geography, provide approach propositions and prolific modeling hypotheses. In ecology, for example, one of the questions most often addressed is the comprehension of the specific diversity in one place and at a given moment: why are there so many species of snails in the Appalachians[1] [PIM 99]? This question is linked to community ecology where we are attempting to understand the interactions between species. A community is a set of species living at the same moment at the same place. There is no need to interact between species of the same community. A traditional question of community ecology, ever since the discipline was revived in the 1960s, particularly by Hutchinson [HUT 59], MacArthur [MAC 67] and others,

---

[1] The diversity of snails in the Appalachians is very large, as with amphibians, but not with reptiles. The question of why this is remains open today.

is to understand the diversity of the communities from interactions between their individuals. As with a network of cities, a community can be represented by a vector whose coordinates are the resources of its different species [SAN 84].

9.2. Formalizations

We will briefly present in this section the most traditional and current ways to model spatio-temporal dynamics. Strictly speaking, spatio-temporal dynamics is a dynamical system representing the evolution of a unit system in interaction, in which these interactions are driven by the spatial organization of the units. Space then acts as a parameter of the system. First, we will limit ourselves to the systems for which the units have a set position in space, such as city systems or plant communities in vegetable ecology, for example. The traditional formalizations[2] distinguish three mathematical spaces:

– space U, or space of phases, which contains the state variables (size of a city, biomass of a plant, for example), in other words, such that a vector u of U describes the state of a unit of the system;

– time T;

– space X within which the units are located.

Each unit k is therefore characterized by its state $u \in U$, time $t \in T$ of observation, location $(x,y) \in X$ of the unit. We observe that each of these variables can be continuous or discrete:

– if d variables describe the state of a unit of the system, $U = \mathbb{R}^d$ or $U = \mathbb{Z}^d$;

– time can be measured in a continuous way $T = \mathbb{R}$ or in a discrete way $T = \mathbb{Z}$;

– space can be continuous $X = \mathbb{R}^2$ or discrete $X = \mathbb{Z}^2$.

The combination of these different possibilities theoretically leads to eight possibilities depending on whether state, time or space is discrete or continuous. Among them, only four are common in the modeling of natural phenomena, and are presented in Table 9.1.

---

[2] There are other possible formalizations, especially in statistical mechanics where the space variables are state variables. These are not developed here because they are not common in geography or in ecology.

Space	Time	Variable	Name
Continuous	Continuous	Continuous	Partial differential equations
Discrete	Continuous	Continuous	Coupled differential equations
Discrete	Discrete	Continuous	Coupled maps on a lattice
Discrete	Discrete	Discrete	Network of cellular automata

Table 9.1. *The main mathematical types of spatio-temporal modeling*

This formalization is common in almost all the application fields of mathematical modeling [JAC 91], in particular the turbulence in physics [FRI 95], which is the preferred field for the study of spatial structures. A question naturally emerges. Do the result of modeling depend on the formalism chosen? In other words, does cellular automata type modeling and partial differential equations applied to a phenomenon lead to the same trajectories? The answer is subtle. Mathematically, it is negative. A well known example is the logistic equation. We know its version in continuous time:

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right) \qquad [9.1]$$

where:

– x: size of the city or biomass of the plant, if we use the examples mentioned previously;

– r: speed of colonization (this colonization is equal to the growth rate when x is low);

– K: capacity of the environment.

We also know the discrete "equivalent"[3]:

$$x_{t+1} = \lambda x_t \left(1 - x_{t+1}\right) \qquad [9.2]$$

where λ is a parameter that plays the role of r in the previous equation.

The trajectories of the continuous equation converge to a set and stable point, whereas the discrete system has played the role of "drosophila"[4] to highlight and study chaotic behaviors in discrete systems. However, the characteristic of convergence toward equilibrium or strange attractor in chaotic regime should be retained when we go from a discrete formalism to a continuous formalism in the same system. The behavior of the formalized system should not depend on this formalization. Generally, a discrete time dynamical system on a unidimensional space can be chaotic, whereas it is impossible (in dimension two as well) in continuous time. This explains clearly enough that both discrete and continuous equations are not equivalent.

However, this question is not only mathematical because modeling includes links with data and observations. This contact with reality is provided by measures which are always discrete. They are continuous numbers with a finite precision at a given moment. We can say that a discrete approach is more natural than a continuous approach. In reality, the correct definition of a continuous set, that of real numbers, has only emerged in mathematics in the 19th century. On the other hand, there are often good general results on the behavior of continuous systems. Everything is done as if the necessary conceptualization effort to control the mathematical tool enabled more powerful results. For example, we have very few theorems on the behavior of cellular automata networks, whereas the equivalent continuous theory, that of equations with partial derivatives is highly developed and rich with results. The transition to a continuous description is a mathematical act: it is a transition at the limit, when the elementary step of space, time or measure of the variable tends toward zero. It can be legitimate to impose, for the transition at the limit, the constraint that the behavior of the continuous system is qualitatively identical to the discrete system. If this point of view is adopted, the question drops because it becomes redundant. However, the question can be reversed: the measure is an access to reality but not reality itself. There can be a certain legitimacy in directly representing models in continuous formalism. Since the measure plays a significant

---

[3] Both systems are not mathematically equivalent or, in other words, the continuous equation is not equivalent to the discrete equation when the step in time becomes very small, but this comment is not important for the rest of this example.

[4] We call drosophila an object on which are focused generic studies concerning a given field and largely exceeding the chosen object. Genetics, for example, has developed tremendously by studying the chromosomes of the drosophila, more easily observable than on other species.

part in this discussion, we will address it in the next section with the notion of perfect aggregation.

9.3. Notion of perfect aggregation of variables

Most of the natural objects studied in modeling have a structure identified from the initial stages of modeling. Often, this structure is simplified in a network of hierarchies which are often formalized as nested scales. This is the case in particular with biology, from molecules to ecosystems, but also with Earth sciences, from chemistry to the continents. Knowing how to communicate between these scales or, in other words, make a given result relevant at a scale where it has not been built is a constant concern and a tremendous problem for those who model natural phenomena. Achieving this transfer of scale is the goal of modeling in ecology according to Pimm [PIM 91]. This author notes that the ecological observations that supply the models are very often collected from several sample plots which total a fraction of a hectare, for a few species (often a taxonomic group, as in birds, reptiles, plants, etc.) and for a period limited to three years[5]. From this data, models are built which must respond to questions formulated at landscape scale (several dozen km^2) for a whole community (hundreds or thousands of species), for a few decades, which is common in planning. A model implements a transfer of scale between the scale of information and that of the questions[6]. The same problem is currently raised about when to integrate functional photosynthesis models which operate at cellular scale in models of general atmospheric circulation with grids of 300×300 km [VAN 97]. This concern then returns to the role of space because it is often the large scale spatial interactions that will impede the possibility of simply building a small scale aggregated model (the term scale is meant here in its geographical and mathematical sense). The explicit consideration of space will be addressed in the next section.

Let us consider a system modeled by a dynamical system. We assume that this system is made up of n interacting components, numbered by $i = 1, \ldots, n$. We note by x_i the real variable describing the i-th component and the global system is represented by variable $x = (x_1, \ldots, x_n)$. We suppose that its evolution is given by the dynamical system:

$$\frac{dx}{dt} = u(x) \qquad [9.3]$$

---

[5] Pimm even specified “the period of a thesis”…

[6] This brings us to ask questions in return on the validation of models: the scale of validation is not the one of data collection.

For a reason that should be obvious in a few moments, we will call this system a microsystem and the state represented by variable x a microstate. Let us now assume that we are interested in an aggregated size for a simplified description of the microsystem, like the mean if it exists:

$$\bar{x} = \frac{1}{n} \sum_i xi \tag{9.4}$$

This is a simplified description because two different microsystems can have the same averages. The operation that calculates the aggregated variable will be called aggregation, the aggregated variable will be called macrostate and the system represented by the aggregated variable will be called macrosystem.

We can imagine for a moment that we do not know that the macrostate comes from an aggregated microstate and that we only have access to the observation of a macrosystem. For example, we will assume that the system is a forest where each tree is known by its height. The microsystem is the data of individual heights, whereas the macrosystem is the mean height. It is the variable that we naturally use when we speak of height of a forest: in general we do not specify the height of each tree. Can we predict its evolution or, in other words, can we model it? To be more precise, we start with macrostate $\bar{x}(0)$. In reality, it comes from a microstate $x(0)$ by aggregation. The dynamics of the microsystem is driven by equation [9.1], which will calculate a microstate at moment t by $x(t) = U(t, x_0)$. Function U, which is a solution for the differential system [9.1], corresponds to the mathematical notion of flow [ARN 74]. We can then aggregate the microstate x at time t to get $\bar{x}(t)$. We understand that this value depends on microstate x at initial time, which was "hidden" behind the aggregated macrostate. If different microstates are aggregated by the same average, there is nothing to prove that it will be the same throughout the life cycle of the system. Knowing only the macrostate is generally not enough to calculate its evolution because we have:

$$\bar{x}(t) = \frac{1}{n} \sum_i x_i(t) = f(t, x_0) \tag{9.5}$$

where f is a function of t and of microstate x_0, that we can calculate knowing U, and the macrostate at time t depends on the microstate at initial time and not on the macrostate at this same time.

In this context, we call a *perfect aggregation* an aggregation where, if two microstates are aggregated at a moment by the same macrostate (for example, they have the same mean), it will be the same way throughout the life time of the system.

In other words, knowing only the macrostate at a given moment is enough to model its evolution for all subsequent moments.

This notion appeared in control theory and then was distributed in economy and ecology in the 1980s, particularly from the publications of [IWA 87]. In forest modeling, for example, it is a very common operation that is called modeling at population scale: the forester only models the evolution of the dominant height of the population, independently of the fluctuations of the individual heights of this value. The dominant height is a slightly less intuitive notion than the mean height: it is the height of the one hundred largest trees of the hectare and empirically it is easier to model the dominant height than the mean height. By this choice, the forester does not know the horizontal structure of heights, i.e. its spatial heterogenity which we can explain with a map. By experience, this simplifying operation is possible as long as the internal heterogenity of the population remains low or controlled. Thus, in a humid tropical forest, where this spatial heterogenity is strong, the concept of dominant height is ineffective and rarely used. We can find a presentation of the aggregation technique and indications on its use in silvics in [FRA 00], [KOK 06] and [PIC 99].

9.4. Mean field

The notion of mean field is very old in physics since it dates back to the modeling of the magnetic properties of solids by Weiss in 1908. It is a very powerful approximation tool which often makes it possible to analytically conduct approximate calculations when it is impossible in an exact system. It is the same as achieving an aggregation on interactions with neighbors. Specifically, a spatially explicit system is a system where the network of relations of each individual is explained and taken into account. In a mean field model, we simplify this network by making each individual interact with a "virtual" individual which is the mean individual of the population. This simplification is an aggregation because we replace the individual description of the relations with an average relation. Before presenting and developing this idea, we will present some formalism.

A dynamical system will be spatialized when components $i = 1, ..., n$ are localized in space. A convenient way to represent this is to allocate a position q_i to the position of component *i* or, in other words, to make it clearer:

$$\frac{dx_i}{dt} = u_i\left(x_1, ..., x_n, q_1, ..., q_n\right) \qquad [9.6]$$

Typically, the interactions between two cities i and j depend on their respective populations x_i and x_j, but also on the distance $\left\| q_i - q_j \right\|$ separating them. The coordinates in space play the role of parameter in the dynamical system. If we write $d_{ij} = \left\| q_i - q_j \right\|$, this system becomes:

$$\frac{dx_i}{dt} = u_i\left(x_1, \ldots, x_n, d_{i1}, \ldots, d_{in}\right) \qquad [9.7]$$

It is not common in a system for all the components to be mutually interacting. For each component, we can define its neighborhood, which is the list of the other components influencing its evolution. Usually, a neighborhood is defined by a distance. This can be more complex, however, as in networks of cities where the range of interactions depends on the resources of the cities concerned. We then assume that, for each component i, we know its neighborhood $v(i)$. We then note by $x_{v(i)}$ the vector of values of variable x for the neighborhood of component i and by $\delta_{v(i)}$ the distances to i of the components neighboring i. We then have:

$$\frac{dx_i}{dt} = u_i\left(x_{v(i)}, \delta_{v(i)}\right) \qquad [9.8]$$

Such a system is generally impossible to solve analytically when the function u_i is non-linear.

We then make the following simple hypothesis: we will aggregate the exact neighborhood $\left(x_{v(i)}, \delta_{v(i)}\right)$ of i in an average neighborhood. For example, if the system is a forest, we will make each tree interact with a mean tree at mean distance, and if the considered system is a city system, we will make each city i interact with an average virtual city at mean distance. This approximation is called *approximation of the mean field* and the calculations are developed later in the example of the Ising model, which comes from physics.

In general, the approximate system does not have the same dynamics of the exact system. It is not really an aggregation because the approximate system and the exact system are both defined by the same variable $x = \left(x_1, \ldots, x_n\right)$. The simplification only involves the interactions: the exact neighborhood is replaced by an average neighborhood. There is no perfect aggregation on the set of neighborhoods.

9.5. Example of the Ising model

The Ising model is an old model (it was proposed in 1925 by the Austrian physicist Lenz as a subject of study to his student Ising) that has been widely used to study magnetism in solids. It can be used outside this field and has been used recently in hydrodynamics to study what we is called lattice gas [WEI 89] and in ecology to study the forest dynamics [FRA 95, WIS 92]. It is the modeling of a solid such as a crystal-like lattice of small magnets, each node of the network carrying a small magnet. This magnet can be directed in two directions that we will call here north and south, or *N* and *S*. There is coupling between neighboring nodes explained as: the way that a magnet is directed toward *N* or south depends on an external field and on the direction of its neighbors. It is at this level that the interaction exists. We then write for each node of the network, a probability that it is directed toward *N*, for example, knowing the external magnetic field and the direction of the neighbors. If we choose, for simplification purposes, a unidimensional network, each cell *k* has two neighbors $k-1$ and $k+1$ and we call s_k the direction in *k*. If *h* is defined as the external field, we then define:

$$p\left(s_k = N \middle| h, s_{k-1}, s_{k+1}\right) \qquad [9.9]$$

which is the probability that the direction of *k* will be *N* (north), knowing the external field *h* and the directions of both neighbors $k-1$ and $k+1$. For this, we should attribute, for example, value $s_k = +1$ if $s_k = N$, and $s_k = -1$ if $s_k = \text{S}$. We define interaction energy as:

$$H_k = -s_k h - J s_k \left(s_{k-1} + s_{k+1}\right)$$

where *J* is a coupling constant[7]. We define, in a general way, a global energy for the whole system:

$$H = \sum_k H_k = -J \sum_{\langle i,j \rangle} s_i s_j - h \sum_i s_i \qquad [9.10]$$

where the sum $\sum_{\langle i,j \rangle}$ means “sum on all neighbor pairs”.

---

[7] Indeed, *J* is in factors of coupling terms coming from both orientations.

The basic hypothesis of statistical physics is to propose that if $\omega = (s_1, ..., s_n)$ is a configuration, or a microstate, the probability of its observation is given by:

$$p(\omega) \approx \exp - \beta H_\omega \qquad [9.11]$$

where ≈ means "is proportional to". We then verify that the proportionality constant is equal to $1/Z$ where $Z = \sum_\omega \exp - \beta H_\omega$, the sum being extended over all microstates ω. We should mention that the state ω of the network, for example a network of ten magnets, is a series of the type *NNSNSSNNSS*. The resolution of the network is the calculation of probabilities of these states and the problem comes from the coupling between neighboring cells through the terms $Js_k s_{k\pm1}$. The calculation is quite simple in one dimension and was conducted by Ising, but is very difficult in two dimensions. It has been achieved by the Norwegian physicist Onsager in 1945, in the case where $h = 0$. The case of three or two dimensions with external field $h \neq 0$ is still not resolved, although we have very high quality approximations at our disposal.

We are thus facing a spatially explicit model – because the direction of a magnet depends on the direction of its neighbors – which is extremely simple in its formulation, but in which the calculations are horribly complex as soon as the dimension of space gets closer to the dimension of our surrounding space. We will see in the next section that the mean field approximation helps us resolve this problem in an approximate way.

The trick is not to make a magnet interact with its actual neighborhood, but with a virtual average neighborhood. We define the mean state in this way:

$$m = \overline{s} = \frac{1}{n} \sum_k s_k \qquad [9.12]$$

Let us note by $p_\uparrow$ the probability that a magnet randomly chosen is in position ↑ and in the same way $p_\downarrow$ for it to be in position ↓. We verify that:

$$m = p_\uparrow \times (+1) + p_\downarrow \times (-1) = p_\uparrow - p_\downarrow = 2p_\uparrow - 1 \qquad [9.13]$$

Now we will make the hypothesis that $h = 0$. We then obtain for the Hamiltonian:

$$H = -J\sum_{\langle i,j \rangle} s_i s_j \qquad [9.14]$$

The simplification of the mean field consists of considering each of the neighbors of a cell in state *m*. Let us consider a cell surrounded by *z* neighbors. We will have $p_{\uparrow} \approx \exp - \beta H_i = \exp \beta z J m$ and $p_{\downarrow} \approx \exp \beta H_i = \exp - \beta z J m$, or:

$$p_{\uparrow} = \frac{\exp \beta Jzm}{\exp \beta Jzm + \exp - \beta Jzm} \qquad [9.15]$$

At equilibrium, we have:

$$\frac{m+1}{2} = \frac{\exp \beta Jzm}{\exp \beta Jzm + \exp - \beta Jzm} \qquad [9.16]$$

which, after transformations, becomes:

$$m = \tanh z\beta Jm \qquad [9.17]$$

and is an implicit equation in *m*. The approximation of the mean field, which consists of imagining that each cell is interacting with the neighbors in the mean state *m*, has enabled us to measure this mean state at equilibrium. We have found that in the case of the Ising model, this approximation is qualitatively exact, even if higher level approximations are necessary for better results. We conclude that knowledge of microstates is not necessary for the calculation of the mean magnetization at equilibrium.

9.6. Use of mean field notion in ecology

The notion of mean field has a peculiar history in the consideration of space in modeling. It is very old but seems to have only recently explicitly reached the shores of modeling of natural phenomena, whether in geography or ecology. What is peculiar is the fact that most of the models proposed in these disciplines are mean field type models. This is natural since it is often the only way to evolve in this compromise between the simplicity of the mathematical form and the respect of the highlights of the phenomenon studied, as mentioned in the introduction. The use of the mean field notion therefore has not "revolutionized" these disciplines but has enabled the proposal of a unifying context beyond the objects specifically studied and provided a support for interdisciplinary dialog.

The simplification of the mean field remains an approximation that is sometimes acceptable and sometimes far from the original dynamics. There are more sophisticated techniques to develop approximation toward higher levels, in particular, pair correlations technical developments go far beyond the introductory context of this chapter.

In the rest of this section, we will give some implicit uses of the notion of mean field in ecology of communities and genetics of the populations.

One of the first models to use the mean field notion implicitly is the Lotka-Volterra model (see [HOF 88] for a presentation and a discussion on the mathematical characteristics of this model, or any ecology work such as, for example, [GUR 98]). We can recall the principle: a population of preys, traditionally defined as n and of predators, traditionally defined as p interacting in the modeled way:

$$\begin{cases} \dfrac{dn}{dt} = n(r - ap) \\ \dfrac{dp}{dt} = p(-m + ean) \end{cases} \qquad [9.18]$$

We consider that a prey, in the absence of the predator, evolves by Malthusian growth $\frac{dn}{dt} = rn$ and that predation causes an additional death rate of $-apn$. Similarly, in the absence of prey, the predators starve to death and die at a rate of $\frac{dp}{dt} = -mp$. The presence of preys causes an additional birth rate equal to bpn noted by $eapn$ by presuming $b = ea$, where e is the efficiency of the predation. In reality, the risk of predation is not identical for every predator/prey pair. It is also affected by age, gender, physical form of those two involved, etc. This model is only then a model from average interactions, where behavioral averages are modeled. This is a mean field type model.

A type of mean field model in forestry is the type of individual models independent from distances. They are individual models because the evolution of each individual is followed by an equation of the type $\frac{dh_i}{dt} = \ldots$ if we follow the height and independent of distances because the interaction (competition for the resources such as light, water and nutrients) comes down to an interaction with the mean tree of the population, in the form $\frac{dh_i}{dt} = \phi\left(h_i, \overline{h}\right)$. This methodology is a good illustration of the relative isolation of each discipline and of the possible interdisciplinary horizons if we consider that the typology of forest models was established early in the 1970s [MUN 74], was very explicitly used with the specific vocabulary of forestry and that the convergence with mean field type models was not explicit until recently.

In population genetics, we focus on the diffusion of genetic information within a population, either by the transport of pollen or seeds. It is traditional to model a population by a group of individuals located in space, for example, on a regular grid and to formalize the dispersion of seeds. We traditionally use two types of models: the *stepping stone* model, in which the dispersion is done by diffusion to free neighboring cells, and the “on island” model, in which the seeds produced are mixed in a group that is later dispersed by random rain over the free cells. The on island model is simply a mean field type model of the *stepping stone* model.

9.7. Reaction-diffusion models

The notion of mean field is not the only simplifying notion which enables this compromise between calculation and models; far from it. There is a subset of mathematics where the calculations are understood well and can be successfully conducted analytically: linear algebra or the matrix calculation. We can resolve linear equation systems, linear differential equations, equations with partial linear derivatives, etc. The traditional statistics are largely based on the linear model. The problem of spatialized models comes from the non-linear character of the links between neighbors. Typically, a spatialized system in one dimension with interactions between the two closest neighbors will be written as:

$$\frac{dx_k}{dt} = u\left(x_{k-1}, x_k, x_{k+1}\right) \qquad [9.19]$$

If function u is non-linear, this system is generally untractable. If function u is linear, or in other words if we can say:

$$\frac{dx_k}{dt} = \alpha x_{k-1} + \beta x_k + \gamma x_{k+1} \qquad [9.20]$$

then we can solve the system. Substantial progress can be achieved in the intermediate case, i.e. the case in which only the interaction with the neighbors is linear. We are then looking at the case of reaction-diffusion systems whose study dates back to the works of the English mathematician Turing early in the 1950s. A term of reaction is in the form:

$$\frac{dx_k}{dt} = u\left(x_k\right) \qquad [9.21]$$

and a term of diffusion:

$$\frac{dx_k}{dt} = \mu\left(x_{k-1} - 2x_k + x_{k+1}\right) \qquad [9.22]$$

which comes from $x_{k-1} - 2x_k + x_{k+1} = \left(x_{k+1} - x_k\right) - \left(x_k - x_{k-1}\right) = \Delta_k - \Delta_{k-1}$. If x_k represents the concentration of a product in cell *k*, the term Δ_k represents a gradient of concentration *k* and the term of diffusion $\Delta_k - \Delta_{k-1}$ represents a gradient of a gradient of a concentration, or a Laplacian. In the formalism where space is continuous, we have in fact, for the term of reaction:

$$\frac{\partial u}{\partial t} = f(u) \qquad [9.23]$$

and for the term of diffusion:

$$\frac{\partial u}{\partial t} = \mu \Delta u \qquad [9.24]$$

Finally, we obtain:

$$\frac{\partial u}{\partial t} = f(u) + \mu \Delta u \qquad [9.25]$$

These systems have been studied, and still are today, on a mathematical level [GRI 91]. They are used well beyond biology. We should mention for history's sake that such equations were introduced in biology in 1937 by Ronald Fisher, one of the founding fathers of modern biometrics, but in the eminently questionable context of eugenics [FIS 37], by choosing $f(u)=u(1-u)$. The corresponding equation including the diffusion term is called Fisher equation and models the propagation of a gene in a population. The use of these equations in biology has been extensively developed recently by Murray [MUR 93].

When function f is non-linear, surprising and "counter-intuitive" phenomena can emerge. A situation of local equilibrium of the reaction system, verified in every point x of space, is a fixed point. In fact, the term of reaction is zero since, by definition, it is a fixed point and the term of diffusion is equally zero since there is no gradient for quantity u. Let us now imagine two species $u(x)$ and $v(x)$ interacting and a system of reaction diffusion where the reaction is the interaction between species, and that there is diffusion of the species by random migration in space. If, at each point x, the resources of both species are those of the equilibrium of the reactive system: $u(x) = u^*$ and $v(x) = v^*$, then the terms of diffusion Δu and Δv are zero. The system with diffusion is also in equilibrium. The "counter-intuitive" result is that this equilibrium can be unstable whereas the diffusion tends to smooth the rough spots and consequently go back to spatial homogenity. The equilibrium reached is a mosaic of patches of one species or the other. It is now relatively easy to simulate at equilibrium a pattern defined in advance like, for example, the spots on the skin of an animal. There is also a whole debate on the relevance of this approach, some emphasizing the risk of artifact as long as the equations of the dynamics will not have been interpreted in a functional context, as is the case in chemistry for oscillating reactions.

9.8. Conclusion

In this brief introduction to the mathematical formalization of spatial interactions, we have reminded our readers that, beyond the technical problems of analytical resolution, the consideration of space raises questions that rigorous resolution has not yet reached. Many natural systems, city networks or forest communities, for example, are modeled by dynamics where the driving forces are individual-based. The natural tendency then is to propose individual models that correspond to this scale and that we call: "individual models depending on the distances" in silvics, and "microsimulation" in geography. The risk of such an approach is to limit ourselves to the model at a 1:1 scale which teaches us very little in relation to the knowledge that is put in. Another, more productive modeling route consists of asking the question of the details that are important in order to understand the global dynamics. This approach is done by the aggregation of variables. We have given an example of the mean field where the details on data spatialization are not mandatory to predict the global variables at equilibrium. This example is taken from a traditional and simple model of statistical physics. The relevance of such a simplification for more complex systems remains an open and largely debated question.

9.9. Bibliography

[ARN 74] ARNOLD V.I., *Equations différentielles ordinaires*, Mir Editions, Moscow, 1974.

[BOU 99] BOULEAU N., *Philosophie des mathématiques et de la modélisation – Du chercheur à l'ingénieur*, L'Harmattan, Paris, 1999.

[FIS 37] FISHER R.A., "The wave of advance of advantageous genes", *Annals of Eugenism*, 7, p. 355-369, 1937.

[FOR 92] FORD E.D., SORRENSEN K.A., "Theory and models of inter-plant competition as a spatial process, in Individual-Based Models and Approaches in Ecology", in *Populations, Communities and Ecosystems*, De Angelis D.L., Gross L. (eds.), Chapman & Hall, New York, p. 363-407, 1992.

[FRA 00] FRANC A., GOURLET-FLEURY S., PICARD N., *Introduction à la modélisation des forêts hétérogènes*, ENGREF Editions, Nancy, 2000.

[FRA 95] FRANC A., BESNARD J., KLEIN E., "Simulation de la dynamique des peuplements forestiers hétérogènes: quelques pistes avec des modèles simples", *Rev. For. Fr.*, vol. XLVII, p. 183-194, 1995.

[FRI 95] FRISCH U., *Turbulence*, Cambridge University Press, Great Britain, 1995.

[GRI 91] GRINROD P., *Pattern and Waves – The Theory and Applications of Reaction-Diffusion Equations*, Oxford University Press, Oxford Applied Mathematics and Computing Science Series, 1991.

[GUR 98] GURNEY W.S.C., NISBET R.M., *Ecological Dynamics*, Oxford University Press, 1998.

[HOF 88] HOFBAUER J., SIGMUND K., *The Theory of Evolution and Dynamical Systems*, Cambridge University Press, Great Britain, 1988.

[HUT 59] HUTCHINSON G.E., "Homage to Santa Rosalia: why are there so many kinds of animals?", *The American Naturalist*, vol. 93, p. 145-159, 1959.

[IWA 87] IWASA Y., ANDREASEN V., LEVIN S., "Aggregation in Model Ecosystems. I Perfect Aggregation", *Ecological Modelling*, 37, p. 287-302, 1987.

[JAC 91] JACKSON E.A., *Perspectives of Nonlinear Dynamics,* vol. 1 and vol. 2, Cambridge University Press, Great Britain, 1991.

[KOK 06] KOKKILA, Y., MAKELA, A. and FRANC, A., "Comparison of distance dependent and distance independent stand growth models: is perfect aggregation possible?", *Forest Science*, 52, p. 623-635, 2006.

[MAC 67] MACARTHUR R.H., WILSON E.O., *Island Biogeography*, Princeton University Press, Princeton Monographs in Population Biology, 1967.

[MUR 93] MURRAY J.D., *Mathematical Biology*, Springer-Verlag, Biomathematics Texts vol. 19, Berlin, 1993 (2nd edition).

[PIC 99] PICARD N., Passage d'un modèle individuel à un d'un modèle de distribution de la dynamique forestière – Application à une forêt dense tropicale humide de Guyane francaise, University thesis, Engref and Cirad Forêt, 1999.

[PIM 91] PIMM S.L., *The Equilibrium of Nature*, Chicago University Press, 1991.

[PIM 99] PIMM S.L., "Terrestrial ecoregions of North America: a conservation assessment", *Book Review section*, *Nature*, 402, p. 853-854, 1999.

[RIC 71] RICHARD P., JAULIN R., *Anthropologie et Calcul*, UGE, Coll. 10-18, Paris, 1971.

[SAN 84] SANDERS L., Interaction spatiale et modélisation dynamique. Application aux systèmes urbains, Paris University Thesis VII, 1984.

[VAN 97] VAN GARDINGEN P.R., FOODY G.M., CURRAN P.J., *Scaling Up – From Cell to Landscape*, Cambridge University Press, Collection Society for Experimental Biology vol. 63, 1997.

[WEI 89] WEISBUCH G., *Dynamique des systèmes complexes*, InterEditions/Editions du CNRS, Paris, 1989.

[WIS 92] WISSEL C., "Modelling the mosaic cycle of a Middle European beech forest", *Ecological Modelling*, 63, p. 29-43, 1992.

Chapter 10

Fractals and Geography

10.1. Introduction

Is the craze for fractals only a fad? Over the last two decades, fractals have fascinated us with the images that they have produced or reproduced but, in social sciences applications, they have fallen short of the expectations of researchers. The measures that have been generated by them have not established themselves alongside more traditional indicators and their use in dynamic models of complex systems has proven to be difficult to implement. It is natural though that they still inspire research in geography, in particular spatial analysis. The irregular and fragmented forms of relief or, urban patterns, the ramifications of hydrographic or transport systems, the hierarchized structures of the world's territories and city systems all have properties, and fractal analysis could propose new interpretations. The self-similar morphology of fractal objects, reproducing the same structures at different scales, is an important feature of the spatial organization of several geographical objects. This essential property has been used in explanatory theories of hierarchical systems, as with central place theory for city systems, but with spatial models that were based on traditional geometry. Introducing fractal geometry as a reference in geographical models is therefore a way to demonstrate certain specific processes of the spatial organization, particularly cities and systems and to find new expressions, especially for dynamic interpretations.

Chapter written by Pierre FRANKHAUSER and Denise PUMAIN.

10.2. Fractality and structuring of the geographical space

The spatiality of human societies is a paradox. On the one hand, when societies humanize and space the land's surface [PIN 88], their development generates homogenity on the land's rough surface in terms of farming and lifestyle, for example, but also in a more general way, in terms of circulation conditions, at least within the boundaries of each large territorial political system. On the other hand, even within a given culture, the human implementation creates new disparities, in particular in terms of spatial distribution of the population and activities that increase as urbanization progresses. The center-periphery duality established itself and became more pronounced with the development of cities, as well as with the expansion of the interactions which contribute to the sizing of territories actually interacting. Its cumulative construction results in an increasingly strong hierarchization of the area, depending on its degree of social complexity [PUM 97].

The fundamental heterogenity of the humanized space is discernible at different observation levels, with a reproduction of similar structures at several organization levels. These level nesting effects have often been reported. The self-similarity of the geographical space has been interpreted by Philbrick [PHI 57], for example, as the result of an almost systematic alternation of polarization effects, which differentiate a center and its periphery, and the similarity effects, which define homogenous regions at all geographical scales: homogenous parcels of an operation are polarized by a farm, the union of several operations forms a small homogenous agricultural region polarized by a market town, several agricultural regions form the periphery of a small regional capital, and the capital itself becomes, with others of its kind, a region polarized by a major city, etc. This functionalistic and static interpretation must be enhanced by the explanation of the reach of spatial interactions [GRA 99], as well as of the speed of movements between regions at the different scales and their historically differentiated evolution [BRE 99].

10.2.1. *Density: a traditional but unsuitable measure*

The measure of human presence most widely used by traditional geography is the measure of density. Compared with usual social indicators that report incomes, services, etc. for an individual, this measure is specific to geography since it indicates quantities on a surface and puts the population at the numerator and not at the denominator of the measure. In reality, population density, long considered as the geographical indicator of choice, measures what we would call today the *ecological capacity* of territorialized socio-economic systems. Density is an analog measure, in its construction, to performance indexes. As with hundredweight of wheat per hectare, we count people per square kilometer. In this capacity, the conceptual efficiency of the notion of density is limited to the situations where there is an actual ecological relation between a surface and the population that develops the resources, and between the land and the human mass that it supports. This relation makes sense in the characterization and comparison of agrarian economies. In the case of cities, which by definition do not practice agricultural activities, the connotation of productivity that the density indicator carries does not apply.

Besides, density necessarily refers to a situation of *homogenity*. In physics-chemistry, from which this measure is borrowed, density is a measure particular to a homogenous distribution of particles for organisms in balance in self-contained systems. In geography, it can only help in characterizing relatively homogenous spatial systems, such as, for example, regions where the level of a specific activity would be uniform. By proposing point maps of town population quantities, George [GEO 50] had already criticized the inability of density maps established at this level to explain the heterogenity of the distribution of the population in its adjustments to local surface textures.

The insufficient density measures are completed by disparity indexes of surface quantity distribution conducted on a given pattern of areal subdivisions (a regular quadrat or any irregular grid). Most of these indexes, such as those based on concentration curves, measure the deviations from an equidistribution situation [BRE 96]. This model implicitly assumes a linear type relation of proportionality between population and surface. However, this relation is almost never verified: when it is empirically determined, the relation between population and the surface of units of certain administrative sections often assumes the form of a power law with an exponent that is lower than 1, in general around two-thirds [HAG 73]. In other words, the most populated units are the smaller areas, or the density decreases as the size of the administrative units considered increases. It is due to these systematic variations that the concentration measures give different results depending on the geographical level at which they are measured [ISA 60]. In this respect, Le Bras [LEB 93] rightly denounces the so-called space occupation intensity measures: "80% of the population lives on 20% of the land", which do not specify the aggregation level of the observations (in this case, for example, we would say that a fifth of the most populated French communes hold the four-fifths of the total population). Brunet carries the image to the absurd by emphasizing the arbitrary character and the problem with this notion of space reference by saying that "to the standards of the Parisian subway, the whole world population could be contained in the Territory of Belfort" [BRU 90].

The measures that refer to a homogenous distribution model lose much information by forgetting the systematically heterogenous character of geographical distributions and especially by not integrating the knowledge acquired about the *general form of these disparities, which are always distributed in a geometric progression*. We give two examples relative to urban localizations. At two observation levels, that of the city and that of urban network levels, we measure a fundamental heterogenity of the spatial distribution of urban mass indicators (people, built-up surfaces, activities or flows) in the considered surfaces. At the level of a city, Clark [CLA 51] has shown for a long time that the distribution of resident population densities, as that of the land rents and real estate costs, is organized with a strong gradient decreasing from the center to the periphery, based on a model of exponential or negative power law of the distance from the center. This model, which is still relevant for costs, maintains its descriptive power, even if the competition exerted by tertiary jobs for the more accessible central locations leads to the formation of a central "crater" in the hyperbolic cone, representing population densities in three dimensions. This fundamental model is not called into question by the recent evolution which, due to residential relaxation of the centers on the one hand and of the densification of suburbs and the dispersion of outlying suburbs on the other hand, has considerably decreased the gradients of density distributions in most of the large cities in the world and in Europe, even in small cities with as little as 20,000 residents.

At urban networks level, the territory is unequally occupied by the extremely hierarchized system of cities. There again, the most widely used models for the analysis are not the most adapted because they refer to the notion of uniformity. The measures aiming to test the form of the distribution of the spread of the cities in a territory have used a Poisson distribution as reference [DAC 67] which presumes an equal probability of occupation by cities and does not consider the effects of accumulation characteristic of urban systems. Similarly, the regular hexagon models of Christaller [CHR 33] take into account a hierarchy of the sizes of the cities, but not the disparity of the resulting density. Recently it has been shown that fractal geometry makes it possible to modify Christallers' model, by articulating two spatial systems: one consists of urbanized areas which are concentrated along transportation axes and the other one is a hierarchical axial system of non built-up, rural zones [FRA 05] Finally, the attempts to use spectral analysis to characterize the scale components corresponding to the different levels of the hierarchical organization of urban networks [CAU 85, DAC 67] have come up against the major irregularities of this organization.

10.2.2. *The fractals: references adapted to the space of human societies*

Knowing that the distribution of urban density, regardless of the level considered, is never homogenous, it can be interesting to replace the model of density by a fractal reference that would contain from the outset the information relative to the heterogenity and to the form that it most often takes. Not only would it be possible to directly compare degrees of heterogenity or to integrate this property into models, but we could also hope to discover something new instead of treating hierarchization phenomena as residuals with respect to a homogenous model. Besides, if the space occupation by cities is similar to the images produced by models of reference using fractal geometry, we could then try to understand why it is so by imagining plausible processes that simulate the genesis of such configurations. Plausible in this case means compatible with the urban theory, or rather the urbanistic or socio-economic theories of the formation and growth of the cities. The reference to the fractal model has the advantage over the density model to return more directly to a dynamic conception [PUM 04].

This conceptual evolution moves the center of interest with respect to densities. It is not so much the intensity of the occupation of space that will be considered but its structure, built from an underlying ordering principle, which is represented at different geographic levels and concerns the connection of these levels. This implicit ordering principle can be revealed despite the disparities caused by random fluctuations. For instance, in the hierarchy of cities, the absence of discontinuity between dimensional or functional levels seems instead to be the rule and this contrary to what central place theory predicts. At a finer scale, and a least in what concerns the built-up surfaces, a discontinuity seems to persist between the continuous constructed space of cities and their periphery not yet belonging to the urban cluster [FRA 93]. A recent demonstration of this dual fractal structure of the urban field has been made for European urban areas described with CorineLandcover data on built-up land use by Marianne Guérois [GUE 03 and GUE 08]. These observations could help theoretically justify the use of multifractals by calling for different genetic processes between those that organize the space occupation at the level of the city and those that structure the spatial thread at the territorial scale of city systems.

Finally, the traditional analysis methods place the population or the activities with respect to a space support containing them and whose properties are those of the Euclidean geometry. For a long time now, Harvey has addressed the fruitfulness of the conception of a *relative space* that is defined by the historical and social practice: “It is the activities and the objects themselves which define their spatial field of intervention” [HAR 69, p. 209]. Authors such as Hägerstrand, Cauvin or Muller have declared its non-Euclidean character, deliberately heterogenous and anisotropic, in analyses of space perception, or in the research of cartographic representations more adapted [RIM 86, TOB 79] or still in the research on theoretical geography on the properties of geographical space [BRU 90]. In intra-urban space we have noted that the reference to a homogenous space does not really fit with distributions that are extremely contrasting and organized into very strong gradients, which express and infer simple or multiple, highly polarized center-periphery fields. Similarly, at the level of city systems, we have observed that the geometric models that come from central place theory seem incompatible, due to their reference to a homogenous distribution of population, with the existence of *polarization fields* which are defined by the methods of space occupation by the cities [FRA 93, FRA 96]. Besides, we have been noticing for a long time that the distributions of urban hierarchies were generally well represented by statistical models (Pareto model, still named “rank size distribution” by Zipf, or lognormal distribution studied by Gibrat), which seem compatible with fractal geometry. Thus, the multifractal generators used by Le Bras [LEB 93] to simulate the spatial distribution of demographic growths, which are not explicitly linked in his book to territorial processes, actually present an analogy with a stochastic process of growth distribution, of which we know that it generates lognormal distributions [PUM 97].

In geography, the most important fractal geometry applications involve the morphology of cities ([BAT 94, FRA 93] are only a few of the major studies) or, more broadly, the organization of the population on a territory [ARL 85, LEB 93, LEB 96]. The configuration of transport networks was also the subject of descriptions with the help of fractals [BEN 91, GEN 00]. The work of Dauphiné [DAU 95] is one of the first to have proposed a panorama of applications in geography, including models in physical geography. A recent application in hydrology was presented by Hauchard *et al.* [HAU 99].

First, we will present a few types of fractal objects, as well as the measures used to characterize them, before examining some applications in detail. These applications were most often chosen in urban geography because they might enable us to specify the interpretation of fractal structures in geography.

10.3. Fractal models of spatial structures

Fractal geometry enables us to analyze a spatial structure from a reference other than Euclidian geometry. Different types of “ideal” fractal objects serve as reference for the description of an observed reality. These constructed fractals play a role similar to that of basic figures like the circle, square, etc. in the Euclidian geometry. These objects are differentiated according to criteria inherent to fractal geometry and we introduce specific measures that will characterize these following a fractal logic.

10.3.1. *Surface models*

Figure 10.1 shows several of these theoretical models used for applications in geography. These objects are obtained by repeating a specific operation called generator. The first example is the teragon (see Figure 10.1a). The initial figure is a square the length of $l_0 = L$. The generator operates on the perimeter: each side is replaced by a polyline made up of $N = 8$ elements of length $l_1 = r\,L$, where r is the reduction factor. In the case of the teragon, $r = 1/4$. We verify that the total surface of the object remains invariable. This procedure is repeated for each element of length l_1 during the next step. In this way, the border of the object lengthens, becomes more complex and finally tends toward infinity, whereas the surface remains constant. Such a behavior does not exist in Euclidean geometry, the perimeter having lost the usual characteristics of a line.

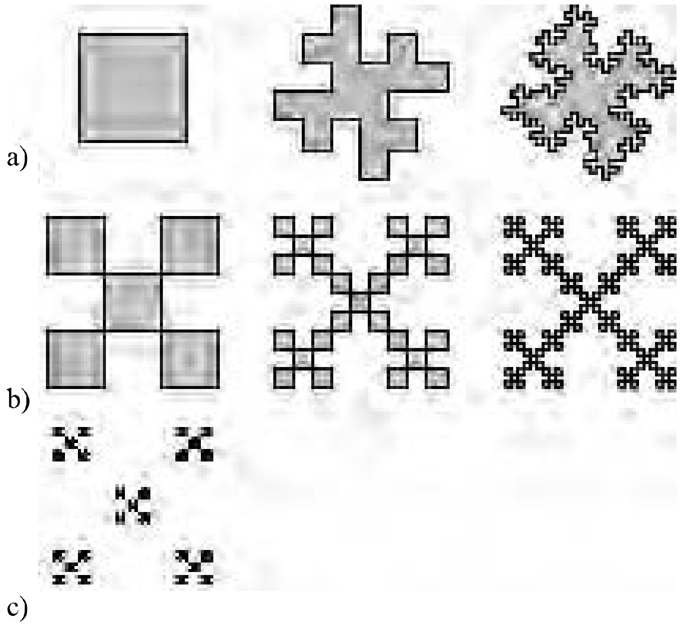


Figure 10.1. *Fractal models of surface occupation*

Two other examples are shown in Figure 10.1. In both cases, the generator reduces the initial figure, a square, by a factor r and we place N of these reduced reproductions, according to a chosen diagram, inside the initial figure. These copies must not intersect. For the Sierpinski carpet (see Figure 10.1b), we obtain a checkerboard where the squares touch each other by their tops, whereas in the Fournier dust (see Figure 10.1c) the squares are isolated. By repeating this operation for each of the N elements generated, we can verify that in both cases the surface of the fractal object approaches zero, whereas the length of the perimeter diverges. The Fournier dust is finally made up of a set of points distributed unevenly: they form masses separated by gaps which are generated during the iteration and whose size decreases at each step. Areas with high concentration and areas of low concentration coexist, whereas a double hierarchy of sizes of surfaces occupied and of the gaps separating them takes shape.

The three examples presented involve the distribution of a mass on a surface. These models can represent the distribution of a type of land use like constructed surface. However, all three structures show differences:

– the teragon remains compact inside, only its border is complex. It could be a model for a city that would remain homogenous (following a certain criterion) inside but penetrating in the space outside the city in a complex way. Such a figure will be mostly useful if we are interested in the perimeter of the built-up space and not in the spatial organization inside the city, for example, to characterize an urban sprawl on the scale of a metropolitan zone, starting from a simplified cartographic representation;

– in the case of the Sierpinski carpet, after a few iterations, only the border effect is significant and characterizes by itself the structure of the phenomenon studied. In fact, border and surface approach the same limit set: a structure in which each point that is part of the fractal set is located in the vicinity of an empty space, but all its elements are related. We could imagine a city that disperses along a divided transport system;

– in the case of the Fournier dust, the aggregates remain isolated. This model can be used as reference to study two types of spatial systems. On an urban district scale, blocks of houses separated by the street network are organized in a hierarchized network and on a regional scale, a distribution of residences with concentrations along the valleys or transport axes.

10.3.2. *Line models*

For certain spatial systems we focus on the morphological properties of linear objects. We can cite, for example, limits such as the edges of a forest or the branches of a ramified system like a transportation network. In the case of the edges, fractal geometry mainly makes it possible to study their complex aspect. Let us recall that one of the examples used by Mandelbrot in the introduction of his book is the analysis of borders done by Richardson. In the network case, two aspects seem interesting: their sinuosity and their junction system which often show a hierarchical organization. It is possible to introduce fractal models to get close to these two notions:

– to represent sinuous networks, we present one of the possible models, the Peano curve (see Figure 10.2a). It is a structure made up of only one line that, due to its sinuosity, tends to cover the whole surface. It represents the model of a route of infinite length, which irrigates the entire surface evenly, which is not the case for other models of this type;

– to represent a ramified network, whose arcs are straight and not sinuous, by reinterpreting the Sierpinski carpet. Let us assume that we are replacing the square by a cross in the generator. By iteration, we obtain a structure becoming increasingly ramified (see Figure 10.2b), called *Sierpinski gasket*. It is organized according to a hierarchical principle with few long branches and an increasing number of branches getting increasingly shorter. This logic recalls a public transportation network concentrating traffic flows toward main axes. In this structure, all the branches are rectilinear and the fractal aspect only appears trough ramifications. The gaps lead to a high local concentration of these branches at each scale and therefore large parts of space are not served by the system. These empty zones form pockets that penetrate in the zones served by the network. In fact, due to the hierarchical organization, there is an increasing number of smaller and smaller zones which are not served

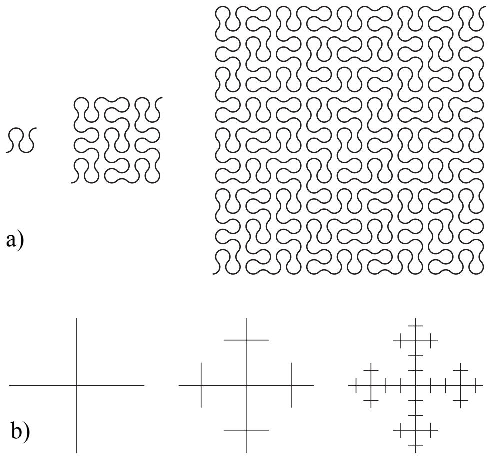


Figure 10.2. *Fractal models of lines and networks*

Finally, it is possible to combine the ramified aspect and the lengthening in one unique fractal model [FRA 93].

The models mentioned above are useful references, but in some ways too simple in their morphological expression with respect to the complexity of the observed structures in reality. To simulate more complex objects, we can modify the value of the reduction factor according to the iterations, either systematically (multifractal models) or randomly and we can also change the position of the generator's elements (stochastic models).

10.3.3. *Multifractal models*

The multifractal approach helps us build more complex fractal models. We obtain a multifractal structure by combining several factors of reduction r_i in the generator. During the iteration steps, we get mixed factors of the type $r_1^n r_2^{n-1} r_3^{n-2}$, etc. The structure is made up of elements of very different sizes.

The multifractal approach is also used to combine a surface padding model, for example, the distribution of the constructed surface from a binary logic of presence-absence of an attribute, and a model introducing a notion of occupation intensity, for example, the local concentration of the population or the height of the buildings. In this case, we affect a different weight p_i to the different generator elements. With the iteration, we get variable combinations of these weights like, for instance, $p_1^n p_2^{n-1} p_3^{n-2}$, etc.

10.3.4. *Stochastic models*

The symmetric aspect of models of regular fractals does not enable comparisons with real structures. In order to obtain figures that seem less artificial, it is possible to introduce random elements in the construction of a fractal (stochastic model), without modifying its fractal properties. For example, in the case of the Sierpinski carpet, since the generator is only defined by parameters N and r, at each step we are free to choose the position of the elements, as long as we respect the gaps generated during the previous steps. We can then get figures that are much more similar to the observed structures. Their mode of construction recalls that of the dynamic models based on cellular automata that are used to simulate the emergence of complex forms (an application by White and Engelen [WHI 94] can be seen in Figure 10.7).

10.4. Measuring fractality

In order to develop a fractal approach for real spatial systems, we must have methods which make it possible to measure the fractal or non-fractal character of these structures. If we try to calculate, for example, the surface of a Sierpinski carpet at each iteration step, we notice that it approaches zero during the iterations, whereas, at the same time, the perimeter tends toward infinity. This fractal is therefore neither a one-dimensional object like a line nor a two-dimensional object like a surface. Similarly, the density proves to be totally inappropriate for the description of distribution of the mass: it largely depends on the position and size of the chosen reference surface.

As a result, the traditional notions of geometry (length, surface, etc.) do not seem appropriate for the quantitative description of fractals. The theory of the measure, however, introduced descriptors characterizing the spatial distribution of the elements, for example the occupied mass, in a fractal. The *fractal dimensions* refer to the main characteristic of fractal geometry, *the hierarchical organization of a spatial system.*

10.4.1. ***Notion of fractal dimension***

The definition of fractal dimensions is directly based on the principle of nested levels. The goal is to analyze the structure by ignoring first of all the details referring to a distance lower than a given size ε. Then, we progressively decrease the size of ε and we thus take into account increasingly fine-grained details. This “zooming” procedure reminds us of the transition of a crude cartographic scaled representation to an increasingly finite resolution. At each step, we count the number of elements $N(\varepsilon)$ (pixels, points, squares, depending on the applications) that constitute the object at this level of resolution.

This approach corresponds to the iterative logic used to generate a constructed fractal, whether it is regular or random. For a Sierpinski carpet, for example, we can interpret the generator as a crude representation of the fractal ignoring the details lower than l_1. The structure is thus composed of $N = 5$ elements (see Figure 10.1b). The next step already takes into account all the details of size $l_2 = 1/3\ l_1$ and the figure is made up of $N^2 = 25$ elements of this size. By following this logic, we discover that for a fractal there exists a constant relation between the number of elements $N(\varepsilon)$ and the resolution ε; a relation which can be expressed in the following way:

$$N(\varepsilon) \sim \varepsilon^D \qquad [10.1]$$

or, in a bilogarithmic way:

$$\log N(\varepsilon) = const + D.\log(\varepsilon) \qquad [10.2]$$

Parameter D, called fractal dimension, is characteristic of this fractal. We can verify that for a constructed fractal we simply obtain, whatever the level of resolution ε considered:

$$D = -\frac{\log N}{\log r} \qquad [10.3]$$

where r is the reduction factor and N the number of elements considered.

In the example given for the Sierpinski carpet, we get $D = 1.49$. If we applied this measure to Euclidian geometric objects, we would get value $D = 2$ for the circle and $D = 1$ for the line. A fractal such as the Sierpinski carpet is located, by its morphology, between a linear structure ($D = 1$) and a homogenous surface ($D = 2$). The value of the fractal dimension characterizes the degree of concentration of the mass in a chosen zone of the structure, in other words, the degree of non-homogenity of the distribution of the mass. A value close to 2 would correspond to a relatively homogenous and therefore slightly hierarchized structure, whereas a value close to zero indicates a strong hierarchy with significant nested concentrations of mass in certain areas.

Several methods of measure are used to characterize the fractal dimension of the structures observed. For constructed regular fractals, we obtain the same value, no matter what method is used. However, for real structures that is not necessarily the case. Observed differences indicate that the fractal behavior varies within the structure, what becomes evident by the ruptures in the analysis curves. This situation reveals a multifractal logic. Multifractal analyses have in fact shown that the range in which the dimensions can vary is an indicator of the complexity. In this case, the use of several methods proves to be enriching because they provide complementary information on one structure.

10.4.2. *Global analysis methods*

The global measure methods give information on the hierarchized organization within a defined zone. In this section, we present two of these methods.

10.4.2.1. *The grid analysis*

The analysis of the grid reminds us of certain procedures in spatial analysis to study, smooth or filter local variations in a group structure in order to reveal scale components. The method consists of covering the structure with a grid for which we vary the dimension of the cell ε. For each value ε, we count the number of cells $N(\varepsilon)$ containing at least one point occupied. By adjusting a relation equivalent to relation [10.2] between $N(\varepsilon)$ and ε, we estimate the corresponding fractal dimension, i.e. the "dimension of the grid" that we note by D_q. This dimension gives general information on the spatial organization inside the zone covered by the grid.

The grid method can be used as a base to conduct multifractal analysis of a structure. In this case, we count the surface actually occupied (1, 2, … or *n* points) that is within each cell. We can then calculate a series of fractal dimensions D_n, $n = 2, 3, 4\ldots$ that indicate the presence of *n* points located at distance ε from each other.

10.4.2.2. *The correlation analysis*

The grid method generates certain ambiguities: the modification of the position or of the size of the window may influence the result. This problem disappears by using another method, the correlation analysis, which turned out to be very reliable. To carry out this analysis, we center a circle of radius ε on each built-up site *k* and we determine for each site the built-up surface $N(\varepsilon_k)$ lying within this circle. Then we calculate the mean value $N(\varepsilon)$ of all these values $N(\varepsilon_k)$. This procedure is repeated for other values ε. When the structure is fractal, a relation corresponding to [10.1] or, respectively, [10.2] is observed. Then a fractal dimension, the correlation dimension, can be estimated. Correlation analysis thus provides information about how globally a chosen zone is structured across scales: if the empirical law $N(\varepsilon)$ is regular, we observe a well developed hierarchical distribution of mass, where there are clusters of non-built-up zones of very different size. If on the graphical representation of this relation we observe a break, we can conclude that spatial organization changes at the corresponding scale, i.e. at a mean distance ε between the built-up sites.

10.4.3. *Local methods of analysis*

Other methods, called local, will either make it possible to analyze a structure from a particular point (for example, the center of a structure) or to emphasize more details in the geographical variations of the fractality.

10.4.3.1. *Radial analysis*

The radial analysis records the information on the spatial organization in the neighborhood of an area called counting center. We surround this point with a circle or a square of size ε which is gradually increased. We obtain a relation equivalent to [10.1] or [10.2], respectively. For a fractal structure, this relation will determine parameter α, the Lipschitz-Hölder exponent, which indicates the "local scaling behavior".

In practice, we can estimate the fractal dimension *D* in two ways: either we can compare the theoretical fractal law [10.1] to the empirically obtained relation $N(\varepsilon)$ and estimate the fractal dimension *D* by means of non-linear regression; or we use a bilogarithmic representation of the empirical relation. According to [10.2] the observed relation should then be linear and we estimate the slope which corresponds to the fractal dimension. Since fractal dimensions are estimated by means of a power law, i.e. a non-linear relation, they are obviously not independent from the models representing density variations. For example, one of the expressions of density $\rho(\varepsilon)$ at distance ε of the center of a city is a negative power function of distance ε at the center. On a two logarithmic coordinates graph, the observations are adjusted by a decreasing straight line of gradient β; β representing the density gradient:

$$\log \rho(\varepsilon) = \log const - \beta \log \varepsilon \qquad [10.4]$$

Batty and Kim [BAT 92] have demonstrated the following relation between the fractal dimension D and the density gradient β:

$$D + \beta = 2 \qquad [10.5]$$

These authors validate the fractal model by emphasizing the fact that the negative power function of the distance to the centre is a better reference model than the exponential function. It has the property of not being modified by a change in scale and can be derived just as “naturally” from a process of entropy maximization (or minimum information). We then have to stipulate that the “constraint” on movements is based on the distance logarithm – which seems more in line with observations relative to space perception than the hypothesis of a linear effect used by Wilson to deduce the exponential model. Hence, we can compare the interpretation of the value of the fractal dimension with that of a density gradient, or of a concentration index, which would remain generally constant on a given surface.

The radial analysis method is well adapted to the study of centered structures. Besides, it is not without analogy with our perception of space. The procedure recalls the situation of an observer positioned in a given place and that gradually widens his field of vision. The logarithmic representation highlights this aspect since it brings out the phenomena close to the counting center, while globalizing the information on distant zones.

10.4.3.2. *The curve of scaling behavior*

Another mode of representation of results enables us to better compare the spatial organization of several structures and to identify the breaks in the fractal behavior. In the bilogarithmic representation, we search for the value of gradient α between each point and the neighboring point. In a theoretical fractal, all these values should be constant and equal to the fractal dimension (i.e. whatever ε, $\alpha = D$). We can represent on a graph the series of these values $\alpha(\varepsilon)$ according to the parameter ε and interpret the gap between the curve thus obtained and a straight line

$\alpha(\varepsilon) = D = const.$ We call the empirical function $\alpha(\varepsilon)$ "a curve of scaling behavior". If the average value of α gradually changes in a certain range of ε, it corresponds to a structural change (at a certain scale) in the spatial organization. On the other hand, fluctuations around the average value of α reveal local deviations from fractal law. If the fluctuations dominate the curve's shape, we may conclude that there exists no particular spatial organization. The range of these deviations becomes an interesting indicator.

10.5. The morphology of contours

Among the first applications of the fractal notion in geography there is the morphological description of continental contours. The coast of Bretagne was chosen, among other examples, by Mandelbrot to demonstrate the ability of fractal geometry to represent natural phenomena. The irregular and fragmented character of the shores can in fact be read with more or less detail depending on the scale of observation. More precisely, the measure of the length of the coast systematically varies with that of the standard that we apply to its contour, whose every decrease leads to the consideration of the details neglected by the measures completed with a longer instrument. In this way, the measure of the length of a coast can become infinite!

A lot of hope has been placed in this model. It was supposed to resolve one of the most irritating problems raised by modern cartography. The old-fashioned, "manual" cartographers could eliminate the useless details from a map that would blur the form of the contours when moved to a smaller scale. Today, who has not felt the visual discomfort caused by computer generated maps, which are admittedly easier for the automatic reproduction of contour lines, but often at the cost of an obstruction along the coastal contours or river lines, overcrowded with useless details or, on the contrary, of excessive geometry of administrative limits. The problem is to determine a degree of generalization of contours that would automatically adapt to the format in which the final map is traced. Several automatic generalization programs using fractal geometry have been proposed for contour smoothing [GOO 97]. Two problems have arisen to oppose their systematic introduction in the geographic information systems (GIS): on the one hand, it is necessary to complete the automatism of the tool with the expertise of the cartographer, who is the only one able to decide if this estuary or that meander is significant on a given scale. On the other hand, it is rare that the contours obey a simple and regular fractality…

The morphology of urban perimeters, generally irregular and mangled, can be caused by the varying dispersion of land available in the periphery; it also depends on the form of the parcels, of the layout of transport networks and of possible policies favoring or not urban density. The fractal measures were applied to this object, particularly by Batty and Longley [BAT 94a], making it possible to characterize periods of generally compact or fragmented urban expansion (see Figure 10.3). Bailly [BAI 96] used expansion methods to analyze the contours of the Mediterranean coastal urbanization in France.

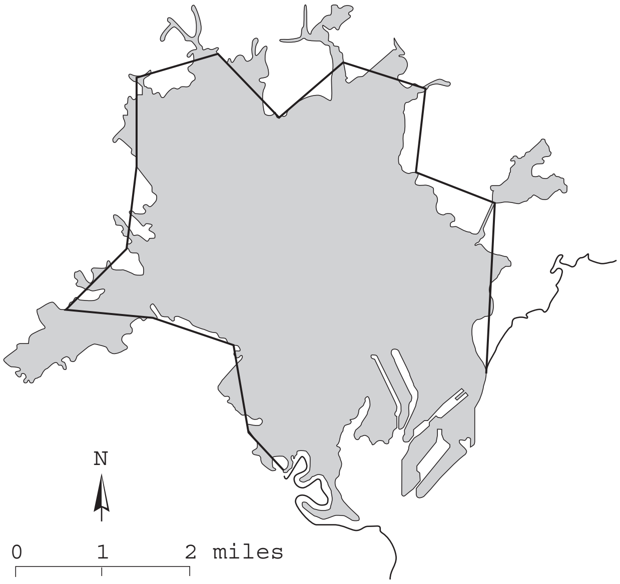


Figure 10.3. *Fractal analysis of the urban perimeter of Cardiff*

10.6. The morphology of land occupation

The fractal surface models are used to characterize the distribution of different kinds of land use like the built-up surface, or wooded spaces, commercial areas, etc. The variable considered is binary since a cell can be occupied or not at each resolution level. When this information of simple surface presence is completed by a measure of the intensity of the occupation, such as a quantity of population, or any census, multifractal reference models must be used.

10.6.1. *Form of occupied surfaces*

The morphology of cities, studied from the distribution on constructed surface, is undoubtedly the application field that has enabled the most in depth analyses and comparisons. Figure 10.4, for example, shows a radial analysis executed from a counting point located in the historical center of Besançon. The bilogarithmic and non-logarithmic graphs (see Figure 10.4a) represent the empirical relation between built-up area $N(\varepsilon)$ according to distance ε to the counting center. On these graphs, we often distinguish distance ranges for which the curves are regular. However, we can detect critical distances ε for which we observe ruptures in the curves. The abscissa of the graphs enables the identification of critical distances ε which correspond to such ruptures. For the example in Figure 10.4, a strong rupture appears in all graphic representation types (Figures 10.4a to 10.4d). For the curve of the scalant behavior (see Figure 10.4c) the rupture becomes particularly evident after smoothing (see Figure 10.4d). The rupture corresponds to a distance of about 350 *m* from the counting point. The corresponding square has been grayed out on the map in Figure 10.4e. It appears that in the case of Besançon, the strong break corresponds to the loop in the Doubs. Hence, the method makes it possible to locate thresholds and discontinuities in an urban texture and to delimit zones according to their spatial organization. When realizing radial analyses with respect to different counting points, it turns out that the ruptures identified often correspond to district limits and often mark the transition from one urbanization period to another.

It is even more surprising to note that the curves describing the scaling behavior of a lot of cities, even if they are of different sizes, show a similar form. Close to the counting center, the scaling behavior translates the nearly homogenous structure of built-up area in historical centers with values of α close to 2. When reaching the limit of this central zone, values α gradually decrease. We then enter into a larger zone corresponding to the suburban areas, with generally regular curves corresponding to a constant scaling behavior but with a weaker value α. The built-up area is here distributed in another way than in the centers, but still follows a fractal law. Finally, a new transitional zone appears on the outside where the slope α gradually weakens and marks the transition to a peripheral zone, less affected by urban growth than the central cluster. Hence, we discover a limit between space where the organization of construction is structured by the presence of the city and the zone where this distribution is more homogenous and becomes non-fractal.

We may conclude that urbanization generates an *internal order* in the urban pattern despite its irregular feature. This impression is confirmed when we compare for a given town the curves of the scaling behavior referring to different dates before smoothing. We often observe a decrease of fluctuations during urbanization [FRA 00]. We have introduced an indicator R_{cor} which makes it possible to measure the degree of fractal order in this spatial system. This indicator compares the deviations of the smoothed curve to the deviations of the average value of the fractal dimension observed. The indicator leans toward *zero* if the fluctuations are more significant than the deviations between the smoothed curve and the average fractal dimension, thus emphasizing the absence of an apparent order in the spatial organization. On the other hand, we get a value close to *one* if the fluctuations disappear.

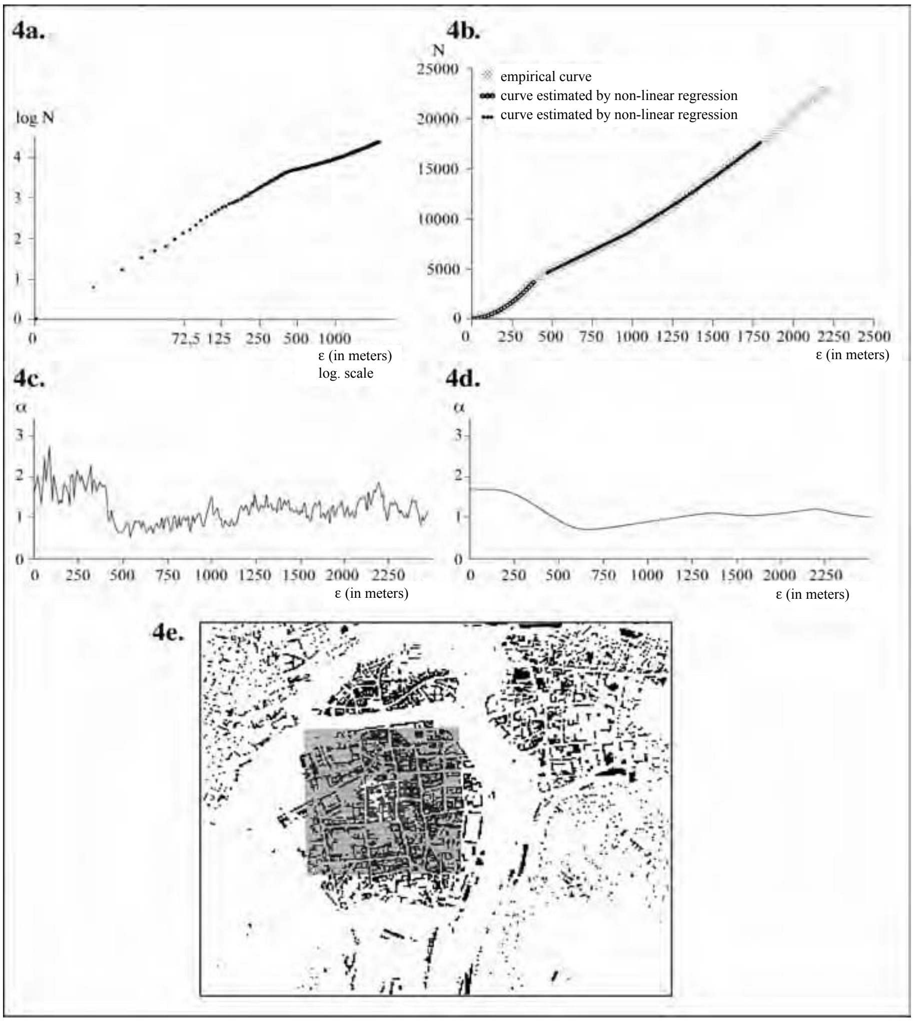


Figure 10.4. *Radial analysis of the constructed surface (center of Besançon)*

Thus, the progressive urbanization of the periphery of large cities generates urban patterns that show a specific coherent organization despite its irregular shape. In the course of urbanization the ruptures in the structures disappear and the suburban villages progressively become a part of a common aggregate. This process is observed for towns of very different sizes, for Berlin and Munich as well as Lons-le-Saulnier, a small town in the east of France (see Figure 10.5). The example of the conurbation of Montbéliard is particular. In the 1950s, a very important Peugeot plant was built: since this period, a significant urbanization has developed close to this site. The degree of order increases for the counting point located in the plant, whereas R_{cor} decreases for the historical center of Montbéliard. The shape of the curve of scaling behavior resembles more and more that of peripheral districts. The city of Audincourt, located further away from the plant, is not as affected by this evolution. The indicator R_{cor}, built like a correlation ratio, plays a role similar to a parameter of order in synergetics.

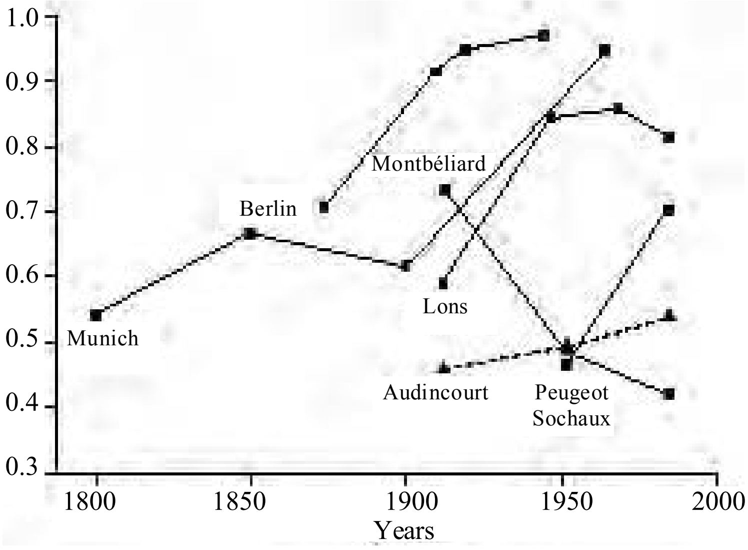


Figure 10.5. *Evolution of the correlation ratio for different cities*

We would expect a lot from the comparisons of the measures of fractal dimensions between cities with similar or different morphologies in order to standardize this new tool. Comparisons have been attempted [BAT 94b, FRA 93, SHE 02] whose analysis remains tricky because the measures are strongly affected by the quality of the basic cartographic documents which, depending on their scale or their degree of generalization, reproduce in a generally uniform way the contours and empty areas of constructed spaces. These measures show the capacity of fractal dimensions to differentiate the spatial organization of urban patterns. Therefore, different values are observed for European towns, which are more structured than

American or Australian towns, where housing is more dispersed in a homogenous way. Comparing a set of European cities has shown that planning policy acts in general on the patterns' morphology of towns [FRA 04]. Moreover recent work which used a coherent database has shown that it is possible to find typical values for the fractal dimension with respect to the historical context of urbanization, the distance to the CBD and other socio-economic indicators like land rent [KEE 03].

10.6.2. *Intensity of land occupation*

The multifractal approach helps us to understand the complexity of the constructed framework, and especially to produce models that simulate a third quantity, for example, population distribution: surface distribution is weighted here by an indicator of population quantity. Le Bras [LEB 93, LEB 96] has created maps reproducing population distribution at a regional scale from databases at NUTS5 level in France and Europe. Population quantities are calculated in this model by the local addition of randomly drawn factors in grids that are very large to start with and then increasingly smaller. A spectrum of fractal dimensions is obtained but it is still difficult to interpret even if the numerous methods of analysis of the population structure, that were previously implemented by the author, have demonstrated a large compatibility with this fractal description.

A model that would be adapted to the description of a population distribution within urban pattern would combine the multifractal logic with that of a Sierpinski carpet. The generator is then made up of empty squares and N occupied squares which symbolize the built-up surface on which the urban population is concentrated. This population is not homogenously distributed; each of the N occupied squares i supports a part p_i. During the next steps of the iteration we generate a hierarchy of empty spaces, but for the occupied surface we obtain weights made up of the combination of factors p_i of the type $p_1^n \, p_2^{n-1} \, p_3^{n-2}$ and consequently a complex distribution of the population.

The analysis of the population distribution at the scale of a city is generally done at the scale of the street block. We then consider that the population is distributed in a homogenous way over the constructed surface of a street block. The multifractal analyses are executed in study zones covered by a grid. Analyses conducted for Besançon show that the range of dimensions D_n, which is determined for each zone, illustrates the degree of complexity of the urban pattern: a low degree of complexity in the downtown area and for the planned city of Planoise, but with a different spatial organization (relatively homogenous for the downtown area and more contrasted in Planoise) on the one hand and, on the other hand a higher complexity with high variations of values D_n for the peripheral suburbs, where the urban texture is more irregular. Figure 10.6 shows an example of this analysis where the gray levels correspond to the intensity of land occupation. The sequence of dimensions refers to the zone of the window shown.

With this type of analysis, we can distinguish different organization types as well as the degree of order in a spatial system. By using data on the height of buildings, it is now possible to measure "unevenness" of the surface of a city, which conditions the microclimate by enabling or stopping the air circulation.

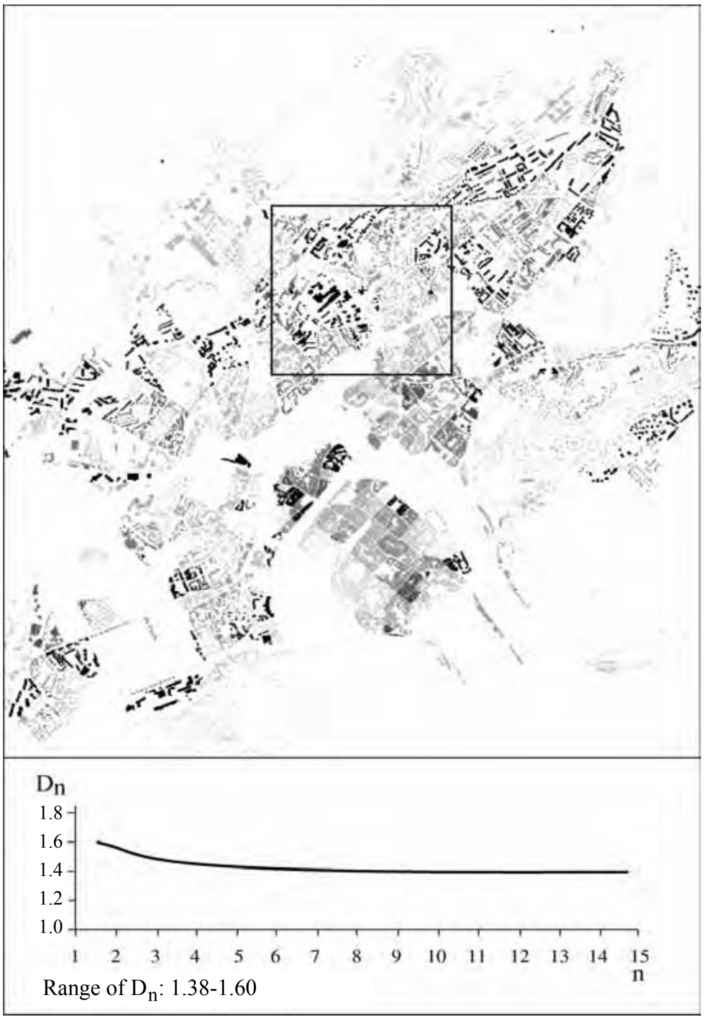


Figure 10.6. *Multifractal analysis of the population distribution of Besançon*

10.7. The morphology of hierarchies: population and systems

The objects that present a strong hierarchical and spatial structure such as city systems, or hydrographic or transport systems lend themselves well to fractal modeling. However, there are not enough applications for these new measures to establish themselves as a complement to those provided by the more traditional methods of analysis for urban systems or for the theory of graphs.

10.7.1. *Urban hierarchies*

One of the first texts discussing the interest of fractals in geography [ARL 85] suggested that geometry of central places was a subset of the theory of fractal geometry and that a fractal iterative sequence, with different generators, could produce all the systems of possible central places. François [FRA 96] has revived the demonstration based on the models of Christaller. In some respects, there is an obvious family tie between the formulation linking the number and size of cities according to Zipf, a Pareto type rule and the fractal model [PUM 06]. However, the application of a radial analysis method to the distribution of French urban units on the territory produces results largely dependent on the position of the counting center [FRA 95]. The measure of fractal dimensions is sensitive to the possible existence of a structure center, role that is played here by the Parisian region, by its exceptional dimension and its structuring role in the system of French cities.

10.7.2. *Measuring the morphology of networks*

Benguigui and Daoud [BEN 91] and Thibaut [THI 96] were the first to conduct fractality measures, respectively on the layout of the Parisian subway and on urban technical networks in Lyon. Generally, it is the radial analysis method that is used to measure the quality of guaranteed service by transport networks. The progressive decline of the service quality from downtown is often well defined by a fractal rule. Ruptures also appear in the periphery of the central business district, similar to those obtained for the built-up areas. The analysis focuses on the spatial distribution of the segments of the network or on the number of stations. If both systems approach the same fractal behavior, the stations are distributed homogenously on the network and their average distance is close to being equal. On the other hand, a fractal dimension that is weaker for the spatial distribution of the stations indicates an increase in the distance between stations when we move away from the center, which is often the case.

The radial analysis gives information that is limited to the morphology of systems. For example, we cannot directly distinguish the part of the fractal behavior resulting from the sinuosity of the network from that due to the ramifications. Other methods are more adapted to the analysis of linear objects. It is this way with the "coastline of Britain analysis", the oldest method for highlighting the extension of a system compared with the Euclidean distance, introduced by Richardson [MAN 82]: we measure the length of a route by replacing it by a sequence of straight line sections of length of ε. By gradually reducing ε, we increasingly take into account the microstructures of the route and we thus obtain information on the progressive lengthening of the route with respect to the Euclidean distance between the starting point and the end point of the route. A similar method, based on Gaussian convolution, smoothes progressively sinuous lines and then allows estimating fractal dimensions.

François and Genre-Grandpierre have used another method which explains the accessibility of sites from a given place [FRA 96, GEN 00]. We start with a digitized network. We then count the total number N_ν of sites accessible in $\nu = 1, 2, 3, \ldots$, steps, from the counting point and by staying in the network. This distance ν is defined as a *connectivity distance* [GOU 92]. In this way, contrary to the radial analysis, we only consider the sites connected at the starting point and the distance of reference, which in this case is not the Euclidean distance but the distance of the network. We then obtain a fractal dimension defined as *dispersion dimension* D_e. It is possible to show that for non-sinuous fractals connected like the Sierpinski carpet, the dispersion dimension is equal to the radial dimension; it is on the other hand inferior to it for sinuous fractals.

Finally, in a sinuous fractal, the connectivity distance ν between the counting center and a point in the network is linked by the following relation to the corresponding Euclidean distance l_ν:

$$\nu = l_\nu^{D_{\min}} \qquad [10.6]$$

We can show that D_{min} is the fractal dimension of the shortest route. In a transport system, it measures the extension of the most direct route. These measures can be used to characterize the distribution of a network's sensitive sections. Besides, it becomes possible to separate the phenomenon of the sinuosity from the ramification of the network, by comparing the results of the radial analysis and those of the dispersion analysis.

Genre-Grandpierre used the analogy of this method with the construction of isochrones to characterize the accessibility of different areas by the dispersion dimension. He has also introduced friction factors which transform the distance on the network in distance-time, from average speeds on the sections of the system based on their equipment level. Real system comparisons at regional level emphasized the modifications, in terms of service, generated by the construction of highway A36 in Franche-Comté. He has also introduced other indicators based on a multiscalar logic describing the quality of average service guaranteed by a transport system. These analyses indicate that the extension of routes plays a more important part in short distance travel and that the current systems implicitly favor long distance travels [GEN 00].

Hence, fractal geometry appears to complement the theory of graphs and can partly complete the limits, especially in the study of transport systems. The first applications are promising but it is imperative that they are improved in greater detail and that the field of the characterization is expanded to other network qualities than only the spatial service if we want fractal geometry to become a real network analysis tool.

10.8. Towards dynamic models

The similarities between observed structures and random fractals have encouraged research for dynamic models for the generation of such structures. Sometimes it is a fractality test that enables us to add an element of plausibility to the simulation result of a spatial structure by a dynamic model, whereas sometimes it is the fractal model itself which inspires the construction of the model. In geography, Tobler [TOB 79] was the first to suggest the interest for this type of model to explain the emergence of complex spatial structures from hypotheses concerning the form of basic spatial interactions. This approach presumes that certain interactions at a microscopic level bring out specific forms at a higher organization level. The microscopic level can be of different nature: it can be districts in interaction or even individuals exchanging information.

In simulations of the use of land, we start with a cartographic representation of the space considered. We then distinguish different categories (residential construction, commercial construction, forest, etc.) and we use a specific cartographic resolution in *raster* form. We assign the use of the land that will be dominating that location at each element of the *raster*, which is defined as a cell. Then we introduce spatial interaction rules between the cells which could locally modify the use of the land. In a traditional cellular automata, we presume that only neighborhood cells influence the dynamics of a site, but in certain models we introduce a longer interaction reach.

The applications can be grouped into two different types of models:

– *the morphological approach*: the authors mainly focus on the generated form. They either resort to diffusion models as they have been used initially in physics (aggregation model by limited diffusion) or to similar models that have, for example, been used to explain electric discharges [BAT 94b]. These models are often modified [FRA 93] for the consideration of particular phenomena, for example, the influence of the surface texture over urban dispersion [BAI 99]. In this case, modeling is already close to the second type;

– *the explanatory approach*: certain authors have introduced in the conception of spatial interaction rules, explanatory factors of spatial dynamics. White and Engelen [WHI 94], for example, define an interaction matrix that takes into consideration the attraction or repulsion effects between certain land occupation types. They presume the negative effect of an industrial zone on the emergence of a residential zone if the zones are too close (annoyances), but a positive effect if the distance is higher than a certain threshold (employee proximity). These authors have used this method to simulate the evolution of the urban growth in Cincinnati (see Figure 10.7).

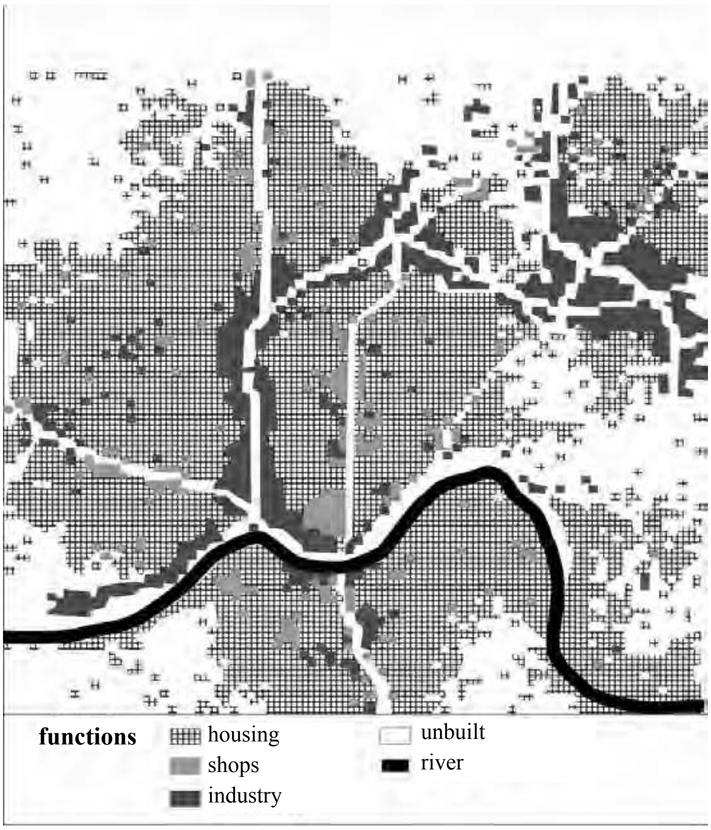


Figure 10.7. *Simulation of urban growth with a fractal model [WHI 94]*

The simulation results approach reality by their morphological appearance. The consideration of spatial interactions is certainly an interesting possibility for testing different hypotheses concerning their influence on the emergence of complex structures. It is also possible to simulate the emergence of urban fabric from simple processes like diffusion. However, the complexity of the models does not always enable us to determine with certainty the real influence of rules on the forms obtained.

10.9. Conclusion

Fractal geometry introduces stimulating references for research in geography, but it is still far from providing solutions to spatial analysis problems. This new type of formalization is still at the stage where the application potential is much richer than the actual developments.

Its contribution in terms of descriptive measures applied to complex morphologies has been until now relatively limited. Too many parameters undoubtedly occur so that the fractal dimension indicators, even if they are very synthetic, are not enough to summarize and differentiate spatial structures. These measures would benefit from being more refined and multiplied, because they are more relevant for spatial distributions, by their reference to a theoretical model, than the usual concentration indices. The finite observation of their local variations is particularly interesting when they reveal thresholds in spatial patterns (for example, and external limit in the urban field, of space structured by urbanization) or spatial discontinuities. These ruptures enable us to differentiate effects of scale which can be explained by a continuity of process (when the fractal dimensions are stable throughout the observation levels) and others, which are explained by different fractalities and, consequently, suggest a modification of generating processes[1].

These static measures only have signification when they are connected with the dynamic processes capable of generating fractal configurations. A large part of research for the simulation of fractal growth was directed by the concern of (re)producing plausible images as Batty and Longley [BAT 94a], for example, have done. Le Bras relies heavily on the resemblance of the images that he obtained with smooth representations of the distribution of the population at county level in France [LEB 93] and in Europe [LEB 96] to discuss the relevance of the fractal model. A method that would transform spatial or social processes into models producing

---

[1] The free ware software package *fractalyse* provides facilities for analyzing the fractal properties of patterns by means of different measuring methods. This software was developed by G. Vuidel under the direction of P. Frankhauser and C. Tannier. The development has been supported by the French Ministry of the Public Works. To download, see www.fractalyse.org.

fractal organizations would undoubtedly be more productive [PUM 98]. We have often demonstrated, particularly with models of hierarchical structures (that of the rank size rule for cities comes to mind), how similar forms could be created from processes of different nature. The study of fractals is still too far from the geographical theory. Frankhauser [FRA 93] proposed some models of urban construction growth simulating concrete mechanisms such as the alternate development of construction and a communications network, or the integration of satellite villages in an expanding urban zone, or still, blocking construction on certain empty zones they can all lead to fractal structures. Research should focus more on the geographical relevance of dynamic processes liable to generate the fractal structures observed, instead of simply trying to reproduce them with ordinary simulation rules. Recent works [FRA 00] have integrated accessibility to different service levels in urban models and measured the degree of optimization provided by the fractal model in terms of space consumption. Moreover, in the work of Cavailhes *et al.* [CAV 04], urban economic modeling was linked to a theoretical fractal model of a town. In this model the household tends to optimize their accessibility to various urban and rural amenities. The obtained results show a complex distribution of land rents which are closer to reality than the usually used radio concentric models.

Fractal models could also be the subject of a prescriptive interpretation when they discover “spontaneous” optimization of certain spatial organizations. In this way, the analysis of human mobilities using transportation means of spatial reach and of different speeds translates into a hierarchization of the systems and a nesting of territories serviced where certain forms of organization could become more efficient or more adaptable than others.

Actual research projects focuses on the potential use of fractal geometry for urban planning. In this context two topics of fractals are of interest: the modified Christaller system makes it possible to develop settlement patterns where urbanized zones and open landscape are mixed. On the one hand, this enables the preservation of large natural areas which are connected to smaller ones localized in the vicinity of urbanized areas. On the other hand, different service levels are introduced. Therefore, the spatial systems simultaneously provide good accessibility to services and leisure areas. By introducing normative reflections, it is then possible to develop a methodolgy which enables the identification of potentials for further development. Thus, it will perhaps be possible to develop instruments directly adapted to the needs of spatial planning, research that would apply in the decision support field in the medium term.

10.10. Bibliography

[ARL 85] ARLINGHAUS S.L., "Fractals take a central place", *Geografiska Annaler*, 67B, p. 83-88, 1985.

[ARL 89] ARLINGHAUS S.L., ARLINGHAUS W.C., "The fractal theory of central place geometry: a diophantine analysis of fractal generators for arbitrary Löschian numbers", *Geographical Analysis*, 21, 2, 1989.

[BAI 96] BAILLY E., "Position de recherche sur une méthode de détermination d'un contour urbain", *Cybergeo*, 10, 6 pages, 1996.

[BAI 99] BAILLY E., Modèle de simulation fractale et croissance urbaine; étude de cas: Nice, Marseille, Genes, PhD Thesis, Nice Sophia-Antipolis University, 1999.

[BAT 86] BATTY M., LONGLEY P., "The fractal simulation of urban structure", *Environment and Planning A*, 18, p. 1143-1179, 1986.

[BAT 89] BATTY M., LONGLEY P., FOTHERINGHAM S., "Urban growth and form: scaling, fractal geometry, and diffusion-limited aggregation", *Environment and Planning A*, 21, p. 1447-1472, 1989.

[BAT 92] BATTY M., KIM K.S., "Form follows function: reformulating urban population density", *Urban Studies*, 29, p. 1043-1070, 1992.

[BAT 94a] BATTY M., LONGLEY P., *Fractal Cities*, Academic Press, London, 1994.

[BAT 94b] BATTY M., XIE Y., "From cells to cities", *Environment and Planning B*, 21, p. 31-48, 1994.

[BAT 05] BATTY M., *Cities and Complexity*, The MIT Press, 2005.

[BEN 91] BENGUIGUI L., DAOUD M., "Is the suburban railway system a fractal?", *Geographical Analysis*, 23, 4, p. 302-368, 1991.

[BRE 96] BRETAGNOLLE A., "Etude des indices de concentration d'une population", *L'Espace Géographique*, 2, p. 135-147, 1996.

[BRE 99] BRETAGNOLLE A., Effets de la contraction espace-temps sur l'évolution des systèmes de villes, PhD Thesis, University Paris I, 1999.

[BRU 90] BRUNET R., *Mondes nouveaux*, Hachette-RECLUS, Géographie Universelle, vol. 1, Paris-Montpellier, 1990.

[BUS 81] BUSSIERE R., STOVALL T., *Systèmes urbains et régionaux à l'état d'équilibre*, C.R.U., Paris, 1981.

[CAU 85] CAUVIN C., REYMOND H., HIRSCH J., *L'espacement des villes*, CNRS, Mémoires et Documents de Géographie, Paris, 1985.

[CAV 04] CAVAILHES 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, p. 1471-1498, 2004.

[CHR 33] CHRISTALLER W., *Die Zentralen Orte in Suddeutschland*, Iena, 1933.

[CLA 51] CLARK C., "Urban population densities", *Journal of the Royal Statistical Society*, A, 114, p. 490-496, 1951.

[DAC 67] DACEY M.F., "A family of density functions for Lösch's measurements on town distribution", in *Spatial analysis*, Berry B.J.L., Marble J. (eds.), p. 479-485, 1967.

[DAU 95] DAUPHINE A., *Chaos, fractales et dynamiques en géographie*, RECLUS, Montpellier, 1995.

[FRA 91] FRANKHAUSER P., "Aspects fractals des structures urbaines", *L'Espace Géographique*, 1, p. 45-69, 1991.

[FRA 93] FRANKHAUSER P., *La fractalité des structures urbaines*, Anthropos, coll. Villes, Paris, 1993.

[FRA 95] FRANÇOIS N., FRANKHAUSER P., PUMAIN D., "Villes, densité et fractalité", *Les Annales de la Recherche Urbaine*, 67, p. 55-64, 1995.

[FRA 96] FRANÇOIS N., Analyse fractale des réseaux urbains, PhD Thesis, University of Paris I, 1996.

[FRA 98] FRANKHAUSER P., "La formalisation fractale des tissus urbains", *Cahiers de Géographie du Québec*, 42, 117, p. 379-398, 1998.

[FRA 98] FRANKHAUSER P., GENRE-GRANPIERRE C., "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, p. 41-78, 1998.

[FRA 00] FRANKHAUSER P., "La fragmentation des espaces urbains et périurbains – Une approche fractale", in *Structure des villes, entreprises et marchés urbains*, L'Harmattan, Emploi, industrie et territoire collection, 2000.

[FRA 04] FRANKHAUSER P., "Comparing the morphology of urban patterns in Europe – a fractal approach, in European cities – insights on outskirts"*, Rapport COST Action 10 Urban Civil Engineering*, Vol. 2, *Structures*, A. Borsdorf and P. Zembri, Brussels, pp. 79-105, 2004.

[FRA 05] FRANKHAUSER P., "La morphologie des tissus urbains et périurbains à travers une lecture fractale"*, Revue Géographique de l'Est*, [3-4]45-160, 2005.

[GEN 00] GENRE-GRANDPIERRE C., Forme et fonctionnement des réseaux de transport: approche fractale et réflexions sur l'aménagement des villes, PhD Thesis, Franche-Comté University, 2000.

[GEO 50] GEORGE P., *Population et peuplement*, INED, Paris, 1950.

[GOO 97] GOODCHILD M.F., QUATTROCHI D.A. (eds.), *Scale in Remote Sensing and GIS*, CRC Press, London, 1997.

[GOU 92] GOUYET J.F., *Physique et structures fractales*, Masson, Paris, 1992.

[GRA 99] GRASLAND C., *Interaction spatiale et morphologie sociale*, HDR, CNRS, UMR Géographie-cités, Paris, 1999.

[GUE 03] GUEROIS M., "Les formes des villes vues du ciel", University of Paris I, PhD Thesis, http://tel.archives-ouvertes.fr/tel-00004303, 2003.

[GUE 08] GUEROIS M., PUMAIN D., "Built up encroachment and the urban field: a comparison of 40 European cities", *Environment and Planning A*, forthcoming, 2008.

[HAG 73] HAGGETT P., *L'analyse spatiale en géographie humaine*, A. Colin, Paris, 1973.

[HAR 69] HARVEY D., *Explanation in Geography*, London, Edward Arnold, 1969.

[HAU 99] HAUCHARD E., DELAHAYE D., FRANKHAUSER P., "Analyse morphologique de thalweg et comportement scalant – application des dimensions non entières à la géographie", *L'Espace géographique*, 3, 1999.

[ISA 60] ISARD W., *Méthodes d'analyse régionale*, Gauthier-Villars, Paris, 1960.

[KEE 03] DE KEERSMAEKER M;-L., FRANKHAUSER P., THOMAS I., "Using fractal dimensions for characterizing intra-urban diversity. The example of Brussels", *Geographical Analysis* 35, pp. 310-328, 2003.

[LEB 93] LE BRAS H., *La planète au village*, Editions de l'Aube, Paris, 1993.

[LEB 96] LE BRAS H., *Le peuplement de l'Europe*, La Documentation française, Paris, 1996.

[MAN 77] MANDELBROT B., *The Fractal Geometry of Nature*, Freeman, San Francisco, 1977.

[PHI 57] PHILBRICK A.K., "Principles of areal functional organisation in regional human geography", *Economic Geography*, p. 229-336, 1957.

[PIN 88] PINCHEMEL P., PINCHEMEL G., *La face de la terre*, A. Colin, Paris, 1988.

[PUM 93] PUMAIN D., "L'espace, le temps et la matérialité des villes", in *Temporalités urbaines*, Lepetit B., Pumain D. (eds.), Anthropos, Paris, p. 135-157, 1993.

[PUM 97] PUMAIN D., "Vers une théorie évolutive des systèmes de villes", *L'Espace Géographique*, 2, p. 119-134, 1997.

[PUM 98] PUMAIN D., "Les modèles d'auto-organisation et le changement urbain", *Cahiers de Géographie du Québec*, 42, p. 117, 349-366, 1998.

[PUM 04] PUMAIN D. "Scaling laws and urban systems", Santa Fe Institute, Working Paper no. 04-02-002, 2004

[PUM 06] PUMAIN D. (ed.), *Hierarchy in Natural and Social Sciences*, Springer, Methodos series 3, 2006.

[RIM 86] RIMBERT S., *Carto-graphies*, Hermes, Paris, 1986.

[SHE 02] SHEN G., "Fractal dimension and fractal growth of urbanized areas", *International Journal of Geographical Information Science* 16(5), 419-437, 2002.

[THI 96] THIBAUT S., Analyse fractale des réseaux d'assainissement, PhD Thesis, INSA, 1996.

[TOB 79] TOBLER W., "Cellular geography", in *Philosophy in Geography*, Gale S., Olsson G. (eds.), Reidel, Dordrecht, p. 279-386, 1979.

[VOI 93] VOIRON-CANICIO C., “Espace, structures et dynamiques régionales: l’arc méditerranéen”, *Analyse spatiale quantitative et appliquée*, no. 33-34, 1993.

[WHI 94] WHITE R., ENGELEN G., “Urban systems dynamics and cellular automata: fractal structures between order and chaos”, *Chaos, Solitions and Fractals*, 4, p. 563-583, 1994.

List of Authors

Sonia CHARDONNEL
PACTE
CNRS
Grenoble
France

François DURAND-DASTÈS
Géographie-cités
Denis Diderot University
Paris
France

Françoise DUREAU
Migrinter
Poitiers University
France

Alain FRANC
BioGeCo
INRA
Bordeaux
France

Pierre FRANKHAUSER
THEMA
Franche-Comté University
Besançon
France

Einar HOLM
Department of Social and Economic Geography
Umeå University
Sweden

Hélène MATHIAN
Géographie-cités
CNRS
Paris
France

Christian MULLON
Ecosystemes d'Upwelling
IRD
Sète
France

Dominique PEETERS
Department of Geography and CORE
Catholic University of Louvain
Louvain-la-Neuve
Belgium

Edith PERRIER
GEODES
IRD
Bondy
France

Marie PIRON
Migrations, mobilités, territoires
IRD
Bondy
France

Denise PUMAIN
Géographie-cités
Paris I University
Institut Universitaire de France
France

Thérèse SAINT-JULIEN
Géographie-cités
Paris I University
France

Lena SANDERS
Géographie-cités
CNRS
Paris
France

Isabelle THOMAS
Department of Geography and CORE
Catholic University of Louvain
Louvain-la-Neuve
Belgium

Jean-Pierre TREUIL
GEODES
IRD
Bondy
France

Christiane WEBER
Image et ville
CNRS
Strasbourg
France

Index