Simulating Sprawl: A Dynamic Entity-Based Approach to Modeling North American Suburban Sprawl Using Cellular Automata and Multi-Agent Systems

Paul Morrison Kevin Torrens

University College London, Department of Geography

Ph.D. thesis, 2004

Abstract

The aim of this research is to develop new spatial models for studying complex urban systems. The models are designed to simulate entity-level dynamics in cities, extending cellular automata and multi-agent systems in a patently spatial fashion. The suitability of the approach for modeling urban systems is demonstrated with an application to sprawling suburban growth in the context of North American cities.

The thesis discusses recent trends in urban simulation, with emphasis on a 'new wave' of approaches to modeling urban systems. The application of that methodology to the study of suburban sprawl is demonstrated through the development of simulations based on the idea of Geographic Automata Systems (GAS). GAS act as a framework for extending automata-based methodologies from the computing sciences with spatial functionality.

Two GAS models are built and discussed in the context of sprawl formation. The first focuses on sprawl as a growth-based phenomenon, simulating the geographic mechanisms that give rise to sprawl in hypothetical and actual metropolitan regions. The second model approaches the idea from a community-level standpoint, simulating dynamics within a residential submarket hypothesized to be in a sprawling urban area.

The results of the research demonstrate the applicability of the modeling framework for simulating urban systems across a variety of scales. The models also reveal several insights regarding the geographical nature of sprawl in a North American context.

Table of contents

List of figures

List of Tables

For Bert and Juicy. I love you so very much.

Acknowledgements

Many many thanks, first and foremost, to my unofficial advisor, mentor, and friend—Itzhak Benenson—and to his family.

Thanks to my family for all the important things.

Thanks, particularly, to Paul Waddell, who got this Ph.D. started and got me interested in urban modeling in the first place. Thanks also to Marina Alberti, for encouraging my interests in complexity and cellular automata, and Anne Vernez-Moudon, who got me interested in sprawl.

Many thanks to my advisors at University College London—Mike Batty and Paul Longley—for encouraging my interests for so long, for reading my stuff for so many years, for writing some great papers and books that had me interested in this before I even met them, and for introducing me to so many people and new ideas. Thanks, in particular to Paul, for coaching me through my job search and for encouraging me when I needed it the most.

Thanks also to my colleagues at the Centre for Advanced Spatial Analysis for their friendship. My special thanks go to Carolina Tobón and Muki Haklay. You both know why. Particular thanks go to the "crowd down the back", some of the UK's greatest minds—Naru Shiode, Martin Dodge, Yoda, a monkey, and some mummies. Thanks, Naru and Martin, for being such great friends. Special thanks, also, to the brains of CASA Prime—Daryl Lloyd and Ian Greatbatch. Thanks to David O'Sullivan for all the great chats, both academic and non-academic, that we have had over the years. Thanks also to the rest of my friends in CASA for their friendship, some really fun nights out, some fantastic idea-bouncing sessions, and the occasional white Americano from Caffé Nero—Steve Coast (Vegas baby!), Alasdair Turner, Elena Besussi, Bin Jiang, Stefan Kueppers, Sinesio Alves Junior, Wolfram Grajetzki, Torsten Schietzelt, Joana Almeda Simoes, Aidan Slingsby, David Ashby, John Ward, Ozlem Sahbaz, and Maria Attard. Thanks, in particular to Theodore Zamenopoulos, Katerina Alexiou, and Joana Barros for being so kind to me in September and on lots of other occasions.

A trip to Tel Aviv University was particularly influential in cementing many of the concepts in this thesis; thanks to all the people I met there for taking the time to run through ideas with me, particularly Erez Hetna and Juval Portugali.

Thanks to my colleagues at Indiana University for all their support and friendship. Thanks, in particular, to Nik Heynen, for getting me seriously interested in an academic life, and for being a good friend.

Thanks to Neal Stephenson, William Gibson, and Iain M. Banks for subtle inspiration, and to Boards of Canada, DJ Shadow, Deftones, Linkin Park, Korn, Snow Patrol, Fugazi, Sleater-Kinney, Sebadoh, Cat Power, Graham Coxon, Idlewild, Death Cab for Cutie, That Dog, Antifilm, Nirvana, Sunny Day Real Estate, Coldplay, Incubus, and Mogwai for some great distractions. Thanks in particular to Taproot, Buffy the Vampire Slayer, and Swingers.

Helen Couclelis and Roger White took the time to reply to emails that I sent to them many years ago when I first started reading about automata models, and fostered a very early interest in all of this. Later chats with them were particularly instrumental in giving me ideas. Thank you. At several stages over the last few years, I have had the opportunity to chat with a number of people, and those chats have inspired me in little and large ways. Thanks to Ferdinando Semboloni, Takeshi Arai, Daniel Brown. Yichun Xie, Jochen Albrecht, Catherine Dibble, Paola Rizzi, Michael Wegener, Peter Allen, Dawn Parker, Femke Reitsma, Anthony Townsend, Chris Webster, Sir Peter Hall, David Banister, Alan Penn, Denise Pumain, Rolf Pendall, Michael Kwartler, Paul Patnode, Roy Chan, Ken Snyder, Joe Ferreira, Vicki Elmer, John Landis, Elena Irwin, Paul Box, Ronnie Ramlogan, Stan Metcalfe, Ivan Blelic, David Lemberg, Steven Manson, Mike Worboys, Harry Richardson, Andrew Frank, Kai Nagel, Guy Engelen, Itzhak Omer, Joshua Epstein, Robert Axtell, Rick Riolo, Victor Mesev, Ivor Suen, Silvana Lombardo, Hans Skov-Petersen, Elisabete Alves Da Silva, Tony Moore, Keith Clarke, Gil Pontius Jr., Richard Aspinall, and Richard Harris.

Thanks, also, to Harvey Miller and Tom Cova for making me feel so welcome.

This was fun.

Part one: models and methods

Chapter 1. Introduction

"This was the way the Minds spent their time. They imagined entirely new universes with altered physical laws, and played with them, lived in them and tinkered with them, sometimes setting up the conditions for life, sometimes just letting things run to see if it would arise spontaneously, sometimes arranging things so that life was impossible but other kinds and types of bizarrely fabulous complication were enabled." (Banks 1996, p. 139)

Introduction

This thesis is about two things—models and sprawl. The research described in this document focuses on the development of new forms of simulation technology and methodology, for exploring the geographic dimensions of change in urban systems. The techniques developed to construct these simulations are demonstrated with reference to suburban sprawl. Controlled, computational environments are built, in which artificial cities are constructed, replete with an array of spatially-motivated synthetic entities. The systems driving change in these simulations, and the artificial structures used as building blocks, are designed to mimic the conditions that we understand to give rise to sprawl in the real world. In this way, the models can be used to explore ideas and hypotheses about sprawl, virtually, in a synthetic space.

It is an opportune time to write a thesis on these topics. Urban simulation has undergone somewhat of a phase shift in recent years, catalyzed by significant developments in the geographical sciences and other fields. A 'new wave' of models is being developed, extending the capabilities and functionality of previous generations of simulation technology and opening up the space of possibilities for simulation in new and exciting ways. It is now possible to build complex models of urban systems, specified in a highly dynamic fashion and with reference to the individual entities that comprise those systems. Yet, work in this area is very much in early stages as a field of research and many issues remain to be addressed. In particular, there is much room for developing spatially-explicit threads of development and inquiry.

At the same time, urban systems are evolving, or emerging, in surprising ways. This is particularly true in North America, where the urban geography on the ground has essentially been re-written in the last fifty years. The phenomenon of suburban sprawl is the poster-child for these kinds of changes, a new form of urbanization somewhere between Ebenezer Howard's Garden City and Corbusian ideas of urban form, and yet altogether different—a geography of nowhere, as authors have referred to it (Kunstler 1993). While the implications of unchecked sprawl in North America, extending peripherally around cities in swaths of development that are much lower in density and scattered in nature when compared to European cities, are reasonably well-understood, the factors behind the phenomenon are less so.

Phenomena such as sprawl are characteristic of new challenges facing urban simulation—phenomena that evolve, rapidly and dynamically, across multiple systems. Such phenomena are manifest at many levels of observation and have explanations at many scales of geography, from the regional down to the individual. The work described in this thesis addresses the problem of simulating such systems; in particular, it focuses on modeling their geographies of change. This is at once an exercise in developing new spatial simulation technology and an endeavor to apply that technology to the exploration and explanation of emerging urban phenomena.

Research goals

The goals of this research are twofold. The first aim is the development of spatial models, based on automata principles from the computing sciences, but also considering traditional methodology used in urban simulation. The extension of the automata idea to spatial contexts is a particular focus. This latter point is particularly important because automata tools are generally developed in a non-spatial manner; this has important implications when using them to simulate and study *geographical* systems—to some extent the methodology limits the range of research questions that can be explored.

The second aim is to test ideas and hypotheses about the formation of suburban sprawl in North America. The particular approach undertaken is one of building sprawl-like systems, by synergy from the bottom-up, in a bid to explore how such systems function and form, interactively. In particular, the research documented in this thesis is concerned with exploring the geographic dimensions of sprawl formation, and questions are explored in the context of the entities that factor in sprawling systems, and their spatial influence at local- and more macro-scales of observation.

There are a few motivations behind this. As mentioned, new forms of simulation are being developed for urban studies. For the most part, these methodologies have origins as computing media and in many instances are applied to spatial contexts with little adaptation. Cellular automata and multi-agent systems are particularly popular and actually have several internal advantages as spatial simulation tools. However, there remains an opportunity for infusing geographical principles directly into these methodologies, and it is especially important in the context of using the tools for *geographic* research.

North American sprawl is an excellent test-bed for the development of this methodology. It is highly dynamic in space and time. It manifests on the ground with geographically-significant configurations and composition. Moreover, the proposed causes of sprawl operate on a cross-scale basis, propagating through urban systems from interactions between individuals in space through to regional-scale geographies.

Research into the potential causes of North American sprawl has not been overtly geographical in its focus, although geography does seem to be a very significant component of the phenomenon. Work in this area has, for the most part, been carried out by researchers in fields outside (although related to) geography—public policy, economics, environmental studies, city and regional planning, architecture. Understandably, the role of space and spatial mechanisms in explaining the phenomenon are not of utmost importance in those contexts. Much of the research on American sprawl is tied to the supposed costs of the phenomenon, and this has tended to emphasize the role of factors such as economics, policy, and planning. There is room for a spatial perspective on sprawl, and for exploring the roles that geography plays in the dynamics of the phenomenon. In particular, the research in this thesis looks at questions relating to the geographical mechanisms that distribute sprawl-like development over urban space and the influence of geography on the spatial behavior of sprawl 'agents',

There is also a volume of debate about the characteristics of sprawl in America, with much discussion devoted to arguing about what *is* and *is not* sprawl. Again, there is a contribution that geography can make to this debate. Several of the characteristics of sprawl are largely spatial in nature, and a wide variety of methodologies exist for quantifying sprawl in an empirical manner. Moreover, several of these techniques can be used to validate automata-type models, and this is a thread in the simulation literature that has a great deal of currency.

The key innovations offered in this thesis, then, are as follows. First, a patently spatial framework for micro-level, dynamic simulation is offered—Geographic Automata Systems. This framework builds on automata tools, but extends the concept with uniquely spatial functionality. Specifically, the framework enables the delineation of entities based on their basic geographies (whether they are fixed or not fixed in space); specification of movement rules for geographic entities (and movement can take the form of locomotion and migration); the tracking of entities in space by direct and indirect means, and the ability to relate their location to other entities in the model; and the specification of variable neighborhood geographies at many spatial scales. In the framework, *collections* of geographic entities come together to form a *system*—an integrated collective of interacting entities. Because the system comprises objects of both fixed and non-fixed description, human-environment dynamics can be expressed. The framework is tested through implementation with respect to modeling the formation of sprawl as a growth phenomenon at a regional and metropolitan scale, and very local-scale residential mobility dynamics within a community and residential submarket. In both instances, an entity-based approach is adopted, whereby the building blocks of the simulated system are designed with explicit consideration of their geographic attributes and behaviors. Second, as a sprawl simulation, the modeling exercises are designed to explore potential causes of suburban sprawl and to look at the geographic dimensions of change in sprawling city-systems in a North American context. Very little work has been done, previously, on modeling sprawl directly, as a phenomenon in its own right. Most studies focus on general urbanization, in which sprawl may or may not be mentioned, or focus on *aspects* of sprawl but without considering them explicitly in the context of sprawl phenomena.

Outline of the thesis

The thesis is organized in three parts—models and methods in part one; substantive material in part two; and applications in part three.

Part one focuses on modeling and methodological issues. Chapter 2 focuses on reviewing popular traditional spatial models, largely for the representation of spatial interaction and choice-making, and other concepts that will re-appear throughout the remainder of the text. The intention in Chapter 2 is to critique traditional techniques, as a prelude to the introduction and description of newer methodologies that form the foundation of the modeling exercises described in part three of the thesis. However, several of those traditional concepts will re-surface later in the thesis as components of sprawl models.

Chapter 3 describes the conditions responsible for recent developments in spatial simulation. The discussion in Chapter 3 goes some of the way toward explaining why computing media have come to dominate the research landscape for simulation in recent years. Developments in complexity studies are also reviewed and the topic will feature recurrently throughout the thesis thereafter. Importantly, advances in the geographical sciences, and their implications for developments in urban simulation, are discussed. These developments are significant in supporting urban simulation, as will be demonstrated in validation exercises described in part three of the thesis. Moreover, developments in Geographic Information Science serve as the inspiration for the Geographic Automata Systems described in the text. The discussion of catalyzing developments in Chapter 3 also serves as an extension to the overview of 'traditional' spatial models in Chapter 2.

A thorough overview of automata tools is presented in Chapter 4. The concepts and mechanisms described in that chapter serve as the foundation for the simulation work described in this thesis. The usefulness of automata as urban simulation tools should be considered in the context of the critique offered in

Chapter 2 and the developments discussed in Chapter 3. Three classes of automata are considered: basic (or general) automata, cellular automata, and multi-agent systems. Each has origins as computing media (although in distinctly different contexts), and this is important. As alluded to at the beginning of this chapter, these tools must be modified for geographical use. And this, essentially, is what is done in the research described in this thesis. Each of these automata tools is described as computing media, then as simulation media, before being reviewed in terms of their capacity to support urban simulation. In the latter sections of Chapter 4, cellular automata and multi-agent systems models of urban phenomena are reviewed. The review focuses on landscape and land-use models in the context of cellular automata, as these are the predominant used to which cellular automata tools are put in urban applications. The review highlights some of the functionality of cellular automata tools, again as compared to the 'traditional' methodologies discussed in Chapter 2. The review of multi-agent systems focuses on their use in simulating movement—traffic—as this is the popular application domain for those tools in urban research. It also provides a background for movement rules that will feature later in the thesis. The key difference between cellular automata and multi-agent systems in terms of representing space is significant. Cellular automata are widely-used to simulate land and the mechanisms of spatial process are largely transitive and diffusive in those models. On the other hand, multi-agent systems are popularly used to simulate mobile entities, and the emphasis is on locomotive spatial processes.

The Geographic Automata Systems introduced in Chapter 5 merge both cellular automata and multi-agent systems in a way. The automata concept is re-defined, using space and spatial processes as a unifying analogy and basis for definition. The result is a framework that unites cellular automata and multi-agent systems, but also makes use of the developments in the computing sciences and geographic sciences outlined in Chapter 3 and Chapter 4.

Part two of the thesis focuses on substantive issues relating to sprawl and considers the application of the models and methods discussed in part one to application in understanding sprawl. Chapter 6 focuses on sprawl as a phenomenon—its characteristics, consequences (both positive and negative), and potential causes. As will become clear in the sections of that chapter, sprawl is not a well-understood phenomenon, and there is great potential for using simulation to inform the sprawl debate. Yet, the complex nature of sprawl requires the use of innovative simulation technology in research on the topic. Chapter 7 considers the many components of sprawl discussed in Chapter 6 in terms of simulation. The questions examined in Chapter 7 ask: What aspects of sprawl can actually be simulated? What would a robust model of sprawl look like? What research has been done in this area already?

Part three of the thesis is focused on developing empirical models for studying sprawl. The construction of those models is described in Chapter 8 and Chapter 9, as an exercise in applying the Geographic Automata Systems framework described in Chapter 5 and as artificial laboratories for exploring components of sprawl discussed in part two of the thesis.

The construction of a regional-scale model of sprawl, approached from the perspective of the spatial distribution of growth, is described in Chapter 8. This model is used to simulate the geographical dimensions of sprawl, modeling sprawling entities in a dynamic context at a regional scale of observation. The model is built around a Geographic Automata Systems core and used to develop three simulations—two abstract cities and one simulation of the Midwestern Megalopolis region of the United States. The simulations are designed to test a variety of ideas about the geographic determinants of sprawl: the relative impact of the potential causes outlined in Chapter 6, the role of growth in determining the rate of sprawl in a city-system and its geographic distribution, and the potential for mitigating the effects of sprawl under different regimes of spatial development. In addition, validation exercises are performed, treating model output as a sprawl landscape and evaluating its configuration empirically, using spatial analysis.

Chapter 9 approaches the sprawl phenomenon from a different standpoint. The model discussed in Chapter 9 was developed to examine the dynamics that might take place within a suburban 'cell' from the growth model discussed above. The focus is on dynamics in residential mobility, at very local scales, and the resilience to change that bottom-up interactions can have at community and submarket levels. This resistance to change is characteristic of many suburban communities, where it is often considered in exclusionary terms. The simulated community/submarket is modeled on an entity basis, using a Geographic Automata Systems engine as before but at a very fine scale of observation—that of the individual properties and households within it. This is a test for the Geographic Automata Systems framework, to evaluate its ability to support simulation of a diverse range of systems and to support cross-scale dynamics. It is also a simulation to evaluate very local-scale properties of sprawl, to ask the question, what are the residential dynamics within submarkets in sprawl-like communities? Experiments are performed to look at the impact of local-scale change in the economic, ethnic, and demographic profile of the submarket, from the bottom-up.

The thesis concludes in Chapter 10 with a discussion of the implications of this work and the lessons learned in terms of model-building, its application to the study of urban systems, and the potential factors underlying the geography of suburban sprawl in North American cities.

Chapter 2. An overview of spatial models

"'Sometimes you repeat yourself, man.' 'It's in my nature'." (Gibson 1984, p.80)

Introduction

The purpose of this chapter is to provide a *very general* overview of spatial modeling in urban contexts. The intention is to look at *what* gets simulated in a spatial urban model and, conventionally, what methodologies are used to specify those models. This review is limited to setting the background for some concepts that will appear in subsequent chapters, specifically, modeling of *spatial interaction* and *spatial choice*. Also, the discussion is focused on exploring the weaknesses and limitations of conventional modeling methodologies as a prelude to the introduction of a 'next generation' of models in subsequent chapters.

This chapter introduces several important concepts that will feature in later discussion through the thesis: spatial interaction, exercising decisions in space, the relationship between forces of supply and demand in space, the friction of movement, utility-maximization, and other topics. Particular attention is paid to evaluation of the ability of traditional models to account for spatial processes and behaviors, both dynamically and at varying scales. It will be argued that much of the traditional methodology for urban simulation is relatively weak in its handling of space when compared to 'new wave' approaches discussed in Chapter 3 and Chapter 4. Much of this discussion is resurrected in Chapter 4, where automata tools are judged by the same criteria. The techniques discussed in this chapter, and the issues that are raised, inform the development of Geographic Automata Systems in Chapter 5. Also, and as will be seen in part three of the thesis, it is apparent that there is much room for incorporating the traditional methodologies discussed in this chapter in automata models.

The chapter continues with description of techniques for modeling development and land-use, focusing on profit calculations and the bid-rent approach. Following this, traditional techniques for modeling spatial interaction are considered. The discussion on spatial interaction models prefaces later discussions on the modeling of spatial processes in later chapters of the thesis. An overview of entropy-maximizing models describes advances in that area of research, and has close parallels with later treatment of complexity studies and their relation to urban systems. Following this, the next major section examines techniques for modeling decisions in urban contexts, with emphasis on non-hierarchical and nested logit models. This section, in particular, provides background for the later introduction of a simulation of preference-based dynamics in synthetic housing markets in Chapter 9. Following this, a critique of traditional methodologies in *urban* modeling is presented. Techniques are evaluated in terms of their approach to dynamics and detail, as well as the usability, flexibility, and realism of traditional models. This section serves as a precursor to Chapter 3, which deals with a 'new wave' of techniques for building urban simulations, many of which have been pioneered with specific respect to the sorts of issues raised here.

Development as a profit calculation

Very simple models are often expressed as simple functional statements. This is quite common in terms of modeling land development, for example.

Developers will develop a site if they judge that they can turn a profit. A number of factors weigh in on this profit calculation; models of development generally simulate this calculation. Development decisions are reliant on *profit margins*; these are a function of the trade-off between input costs (the costs of developing a site) and the expected selling price of the development. The actual acquisition price of the site for development is normally valued by the *residual method of valuation.* If a developer is able to acquire a site for a price below its residual value, she can (potentially) turn a profit by developing that site. The residual value may be calculated according to the following equation:

$$V = M - C - P \qquad \text{Eq i}$$

In the equation, V refers to the residual value of a development site (often expressed per annum), M refers to the market value of the finished product, C

refers to the full costs of development (and this will often include the costs of development *at a particular site*, thereby introducing a spatial component), and P refers to the developer's required profit on gross development value (Adams 1994).

A number of additional variables may be added to this equation. There may be a range of input costs associated with land development, including expenses such as land, labor, materials, fixed costs, marketing, the costs of capital, fees, and interest charges (Adams 1994; Bramley et al. 1995). (Interest charges strongly influence development on the supply side—if interest charges go up, the cost of borrowing capital for development rises.) The determination of a profit calculation may be further complicated by the issue of the timing of land acquisition in relation to the sale of the development as this impacts upon the cost of the land and the turnover time of capital (Bramley *et al.* 1995).

The bid-rent framework

The bid-rent framework expands on basic calculation-based modeling, with a more explicit consideration of space and the cost of activity at particular locations (Figure 1).

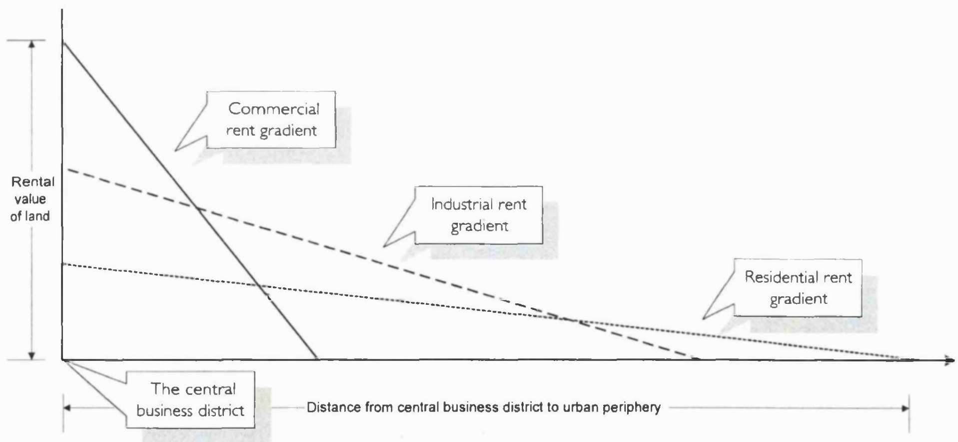


Figure 1. Land-use as described by bid-rent theory

Proceeding from a set of simplifying assumptions (notably, monocentric cities and a limited range of land-uses), bid-rent theory offers an explanation for the spatial distribution of urban activities. Bid-rent theory dates back to very early ideas about cities, essentially to the work of von Thunen (see Carter 1981). It was originally popularized by William Alonso in the 1960s (Alonso 1960), and was later investigated, quite actively, in the context of density gradients (Alperovich & Deutsch 1992; Batty & Kwang 1992; Mills 1972; Mills & Tan 1980; Muth 1969), and also features in 'New Urban Economics' (for example, see Richardson 1977). The central argument of bid-rent theory is that the spatial organization of land-uses is based on whether the land-use is competitive in terms of land rents. That depends on the value, profit, or utility of the land-use and its location in relation to a central urban core. Given these considerations, land-uses tend toward a spatial arrangement akin to that illustrated in Figure 1—with businesses located close to the urban core and industry and residences situated towards the urban periphery.

However, bid-rent theory has a limited theoretical justification in many contemporary urban contexts and does not always transfer to practice easily. The utility function, in particular, can be difficult to calculate. Often, it involves a cost component (e.g., travel time and/or travel cost related to distance from work). In many circumstances utility actually relies on non-monetary conditions that are difficult to cost or measure—the availability of space, fresh air, peace and quiet; location prestige; neighbors; family ties; etc. (Balchin & Kieve 1977).

Similarly, hedonic price models (see DiPasquale & Wheaton 1996) distil real estate values into constituent components (e.g., land value, structure value, number of bedrooms in property, etc.), each of which has a development cost (or consumer value) associated with it. Hedonic price models are quite popular, but are weakened to some extent by their reliance on price; many things are difficult assign monetary value to. Also, because of privacy concerns, price data can be difficult to obtain, especially data spanning multiple time periods.

Bid-rent theory is particularly valuable in many ways, however. Conceptually, it forms the cornerstone of later-developed computational models of urban geographic evolution. It also has relevance in the context of complexity and emergence, which will be discusses in much more detail in later sections of this thesis. At this stage it is worth acknowledging the *emergence* inherent in the bid-rent model; land-uses and urban activities display self-organization under this framework, in a particularly spatial manner, despite the absence, essentially, of spatial mechanisms in the framework.

Spatial interaction

Moving beyond simple functional statements, where space is included as one of many variables, we can consider models of spatial processes. Models of *interaction* are particularly popular. Perhaps the most famous methodology for expressing interactions in space, formally, is the spatial interaction (or 'gravity') model.

Spatial interaction models

The key feature of the spatial interaction model is its representation of interaction as flows between spatially-delineated areas. Spatial interaction models estimate the volume of flows between locations in space, using independent variables that measure some structural properties of the locations considered. For example, the volume of journey-to-work flows might be expressed in relation to structural variables such as the distribution of workers, the distribution of employment, and the costs of traveling to work. The attractiveness of a commercial sub-market for office location might be modeled as a function of floor space, availability of local workforce, etc.

Based on mathematical assumptions that resemble Newton's laws concerning gravitational attraction, gravity models are a particular instance of the broader class of spatial interaction models. Newton asserted that the force of attraction, F_{ab}, between two bodies is the product of their masses, m_a and m_b, divided by the square of the distance between them, d_{ab}^2:

$$F_{ab} = G \cdot \frac{m_a m_b}{d_{ab}^2} \qquad \text{Eq ii}$$

In the equation above, G is a gravitational constant.

Translating this into a geographical context, e.g., intra-urban movement, we could think of flows in terms of migrations:

$$T_{ij} = k \cdot \frac{W_i W_j}{d_{ij}^2} \qquad \text{Eq iii}$$

If we use a geographic example and translate the above equation to a migration model, the task becomes one of allocating migrations in proportion to employment. (An alternative approach to residential mobility will be used later in Chapter 9.) We could assume that people pick their residential location with consideration of their employment location, denoting residential areas with *i* and employment areas with *j*. T_{ij} is the number of people that live in area *i* and work in *j*. The variable for mass, W_i, represents the attractiveness of area *i* as a residential location and W_j corresponds to the attractiveness of *j* as a location for employment, with d_{ij}^2 representing the distance between the two. k is a scaling constant; it needs to be included because the independent variables W_i and W_j are not measured in units of flow (Thomas & Huggett 1980). If we stick to our gravity analogy, it is assumed that the amount of interaction between the two regions, T_{ij}, *declines* in proportion with the square of the distance d_{ij}^2 between the two regions:

$$T_{ij} \propto \frac{1}{d_{ij}^2} \qquad \text{Eq iv}$$

$$\text{or, } T_{ij} \propto d_{ij}^{-2} \qquad \text{Eq v}$$

The validity of this proposition is often justified with data for different types of interaction that show that there is an element of distance-decay in flows such as urban travel, i.e., that short-distance trips occur more frequently than long-distance trips. However, unless we adhere literally to Newtonian principles, there is no theoretical justification for expecting flows to decline exactly with the square of distance between regions. For this reason, it makes more geographical sense to allow distance to be raised to some power α and to rewrite the assumption more generally as:

$$T_{ij} \propto d_{ij}^{-\alpha} \qquad \text{Eq vi}$$

The exact value assigned to α will depend on the available empirical evidence. Raising α to progressively higher powers makes the gradient of the distance-decay curve steep and the value of α is said to measure the *frictional effect of distance*.

Significant variations on this basic description of the gravity model include the production-constrained model, the attraction-constrained model, and the production-attraction-constrained model, which can all be generated using entropy-maximization (this will be discussed shortly). The motivation behind applying these enhancements to the basic framework is to provide some form of balancing or accounting in the predictions that the model makes. Or, put another way, the notion of the constraint serves as a proxy for the theoretical notion of market equilibrium. (Although the inclusion of constraints in the gravity model is perhaps more a function of its weakness in matching observed and predicted flows than of any desire to reconcile the technique with urban economic theory.)

Production constraints

Constraints in gravity models use known information to "fill in the gaps" in calculations. A production-constrained gravity model is one in which the total number of flows leaving an origin i is already known. This knowledge is incorporated into the model design. To recap, let us restate the original gravity model equation and then examine how that formulation changes with the application of a production constraint. The basic gravity model may be formulated as follows:

$$T_{ij} = k \cdot \frac{W_i W_j}{d_{ij}^{\alpha}} \qquad \text{Eq vii}$$

The production-constrained model is a confined version of this formula, where the following constraint is satisfied:

$$\sum_{j}^{n} T_{ij} = O_i \qquad \text{Eq viii}$$

Here, $\sum_{j}^{n}$ sums the values of O_i (usually a value for the population size of a trip or migration origin zone) across all destinations j; T_{ij} is the predicted flow between origin i and destination j; and O_i is the known total number of flows beginning in origin zone i. What the production constraint secures, then, is that the sum of all flows predicted as originating in zone i actually conform to the total number of flows that we *know* originated in that zone. We know how many trips, journeys, or migrations left origins in the urban system in advance of beginning the simulation process, so we can constrain the model to prevent over- or under-predicting of this figure.

Adding the production constraint to the model yields:

$$T_{ij} = A_i O_i W_j c_{ij}^{-\alpha} \qquad \text{Eq ix}$$

where T_{ij} is the predicted flow of trips (or any flow in the urban system) between origin i and destination j, O_i is the total known number of trips beginning in origin i, W_j is the attractiveness of destination j for the flow (e.g., floor space or employment), and $c_{ij}^{-\alpha}$ is cost of travel between i and j with a distance-decay effect applied. A_i has replaced k in the basic model. A_i is a scaling constant for each origin i that ensures that the sum of the flows leaving zone i for destinations j sum to the *known* total zonal flow count. In this sense, A_i is the ratio between the known flow from i and the sum of the unscaled predicted flows leaving origin i for destination j. Mathematically this can be represented as follows:

$$A_i = \frac{O_i}{\sum_j^n O_i W_j c_{ij}^{-\alpha}} = \frac{1}{\sum_j^n w_j c_{ij}^{-\alpha}} \qquad \text{Eq x}$$

Attraction constraints

In an attraction-constrained model, we know how many flows have reached destinations j in an urban system. Again, the predicted flow matrix is made to satisfy a constraint, this time in the form:

$$\sum_i^n T_{ij} = D_j \qquad \text{Eq xi}$$

where D_j is the known number of flows reaching a destination j (e.g., the number of jobs at a destination j, or perhaps the allure of shopping facilities there). Incorporating the constraint fully into our basic gravity model yields the formula:

$$T_{ij} = B_j D_j W_i c_{ij}^{-\alpha} \qquad \text{Eq xii}$$

where D_j is the known number of flows reaching destination j, W_i is the attractiveness of origin i as a source for those flows, and $c_{ij}^{-\alpha}$ is cost of travel between i and j with a distance-decay effect applied. B_j is a scaling constant; for any destination zone j, B_j is calculated as the ratio between the known mass of attractors in that destination (D_j) and the sum of the unscaled flows arriving in destination zone j from each origin zone i (Thomas & Huggett 1980). B_j can be expressed mathematically as:

$$B_j = \frac{D_j}{\sum_j^n D_j W_i c_{ij}^{-\alpha}} = \frac{1}{\sum_j^n W_i c_{ij}^{-\alpha}} \qquad \text{Eq xiii}$$

The attraction-constrained gravity model is commonly used as a residential location model: knowledge of the distribution of jobs, the residential attractiveness of each zone, and journey-to-work costs is used to assign workers to zones in a city.

Production-attraction constraints

The singly-constrained gravity model (either of the production- or attraction-constrained models in isolation) essentially becomes a location model. However, if *both* flow origins and flow destinations are constrained in the model, the emphasis in the model defaults to predicting the size of individual flows (T_{ij}) (Thomas & Huggett 1980). In production-attraction constrained models, the predicted flows are required to satisfy two constraints *simultaneously* (the production and attraction constraints already discussed):

$$\sum_{j}^{n} T_{ij} = O_i \qquad \text{Eq xiv}$$

$$\text{and } \sum_{i}^{n} T_{ij} = D_j \qquad \text{Eq xv}$$

Incorporating these into the basic gravity formula yields:

$$T_{ij} = A_i O_i B_j D_j c_{ij}^{-\alpha}, \qquad \text{Eq xvi}$$

with the scaling properties A_i and B_j appropriately defined as before.

Entropy-maximizing models

The notion of entropy offers a theoretical framework for spatial interaction models (Wilson 1970). Based on statistical mechanics, entropy is concerned with finding the degree of likelihood of the final state of a system. In its simplest definition, entropy is disorder. In the context of cities, data for urban systems are not usually abundantly available. We therefore need a method for making reasoned estimates of the likely state of an urban system using the information that we do know. In this sense, we maximize entropy subject to constraints of known information.

There are two important concepts in entropy that are applied in urban contexts—the macro-state and the micro-state. If we consider our urban system to be comprised of flows between origins and destinations, we may think of the macro-state description of our system as being the numbers of individuals or items flowing between origins and destinations. This macro-state is composed of many micro-states—descriptions of the actual individuals or items that make up a macro-state. Just as there are many possible arrangements of individuals that could make up a subway train of two hundred commuters traveling from one location to another, there are many possible micro-states that can make up a given macro-state.

The number of micro-states associated with any given macro-state can be calculated as:

$$R = \frac{N!}{\prod_{i}^{n} N_i!} \qquad \text{Eq xvii}$$

In the equation above, R is the number of micro-states associated with any given macro-state for the system, N is the number of individuals or items assigned to a set of categories, N_i is the number of individuals in a category i, $N!$ is the factorial value of N: $N : N(N-1)(N-2)(N-3)...1$, and $\prod_{i}^{n}$ is the product of a factorial value.

Framed in this context, the problem of modeling spatial flows then becomes one of maximizing entropy, or disorder—choosing the macro-state associated with the largest number of micro-states (see De la Barra 1989; Fotheringham *et al.* 2000; Fotheringham & O'Kelly 1989; Wilson 1970 for examples). If we consider our flows to be trips from an origin i to a destination j, we can substitute T (the total number of trips made in our system) and T_{ij} (the individual flow of trips from an origin to destination) for N and N_i in the above equation. This yields:

$$R = \frac{T!}{\prod_{ij}^{n} T_{ij}!} \qquad \text{Eq xviii}$$

In dealing with something as complex as an urban system, a modeler can end up with many possible states to pick from in her choice set. As with the basic gravity model, constraints may be introduced into the entropy-maximizing framework, allowing us to reduce the choice set of predicted trip matrices down to a manageable level. One such example is the imposition of a cost constraint on trips:

$$\sum_{i}^{n}\sum_{j}^{n} T_{ij}c_{ij} = C \qquad \text{Eq xix}$$

In the equation above, c_{ij} is the cost of 'flow' from zone *i* to zone *j*, and C is the overall expenditure available for those trips.

Entropy needs to be maximized to arrive at a solution to our problem of identifying the most likely trip or migration matrix from a potentially infinite number of possible forms. The maximization of the entropy value involves the use of Lagrange multipliers (a technique for evaluating maxima or minima of a function subject to one or more equality constraints). Essentially, the Lagrange multipliers serve as weightings to ensure that constraints within the model are met. Incorporating constraints, the model may be expressed mathematically as:

$$L = \ln W + \sum_{i}^{n} \tau_i \left(O_i - \sum_{j}^{n} T_{ij} \right) + \sum_{j}^{n} T_{ij} \left(D_j - \sum_{i}^{n} T_{ij} \right) + \beta \left(C - \sum_{ij}^{n} T_{ij}c_{ij} \right) \qquad \text{Eq xx}$$

where L is the function to be maximized subject to constraints; τ_i is the Lagrange multiplier associated with a production constraint; T_{ij} is the multiplier associated with an attraction constraint (and if both production and attraction constraints are included the model may be considered to be doubly-constrained); and β is the multiplier associated with a cost constraint. The trip matrices that maximize L, i.e., the most likely distributions of trips in the urban area, are solutions of the calculation:

$$\frac{\partial L}{\partial T_{ij}} = 0 \qquad \text{Eq xxi}$$

To solve this equation we make use of Stirling's approximation when the values of T_{ij} are large:

$$\log T_{ij}! = T_{ij} \log T_{ij} - T_{ij} \qquad \text{Eq xxii}$$

We may also maximize $\ln R$ instead of R such that:

$$\frac{\partial L}{\partial T_{ij}} = -\ln T_{ij} - \tau_i - \tau_j - \beta c_{ij} \qquad \text{Eq xxiii}$$

Setting the equation to zero and solving yields:

$$T_{ij} = \exp(-\tau_i - \tau_j - \beta c_{ij}) \qquad \text{Eq xxiv}$$

One of the innovative features of the entropy approach to spatial interaction modeling is that it provides a theoretical rationale (albeit derived from statistical mechanics) for a *family* of spatial interaction models. By substituting the above equation in place of T_{ij} in the constrained gravity models already explored we convert them to entropy-maximizing versions. For the origin constraint, the equivalent entropy model is derived as:

$$\sum_{j}^{n} T_{ij} = O_i \textbf{ becomes } \exp(-\tau_i) = O_i \left[\sum_{j}^{n} \exp(-\tau_j - \beta c_{ij}) \right]^{-1}$$ Eq xxv

and for the destination constraint, the equivalent equation is:

$$\sum_{i}^{n} T_{ij} = D_j \textbf{ becomes } \exp(-\tau_j) = D_j \left[\sum_{i}^{n} \exp(-\tau_i - \beta c_{ij}) \right]^{-1}$$ Eq xxvi

To see how this results in a full spatial interaction model, we simply convert the multiples to out our scaling constants, A_i and B_j:

$$T_{ij} = O_i D_j \exp(-\beta c_{ij}) A_i B_j$$ Eq xxvii

This represents the entropy spatial interaction model in its general form. From that equation, four versions may be derived: origin-constrained, destination-constrained, doubly constrained, and unconstrained.

Choice models

Not all interactions in an urban system occur as flows. At the start of the 1970s, some serious criticisms were leveled against gravity-type formulations of spatial models. In reaction to this, modelers began to develop spatial simulations that were more behaviorally grounded. One avenue of development that was widely embraced was discrete choice modeling. Broadly speaking, discrete choice models are concerned with explaining phenomena in terms of choice-making. While they function in a fashion that resembles spatial *interaction* models, they are actually concerned with spatial *choice*. The most widely used manifestation of the discrete choice model in urban simulation is the random utility model and variations thereof.

Random utility models proceed on a number of assumptions. The first assumption specifies that each choice-maker is faced with a discrete set of choice alternatives—a choice is either made or not made. The second assumes that an individual (or a group of individuals) will settle upon one decision from a larger set of available options in such a way that the most utility, or satisfaction, is yielded. Contextualizing this in an urban sense, we might think of a household making a location choice amongst a set of given locations that a city has to offer so that a combination of utilities is maximized (e.g., cost, amenities, quality of the school system, etc.). The third assumption in random utility models is that choices are made in a probabilistic fashion—choice-makers have a likelihood of making certain choices. Finally, it is assumed that the utility of a decision can be divided into two components: one measuring 'strict utility': the fixed and measurable attributes of utility, and the other dealing with stochastic utility: an error or disturbance term that reflects the unobserved attributes of a given decision (De la Barra 1989; Louviere *et al.* 2000).

Mathematically, we can build up a formula for the random utility model based on these assumptions. The notion of *utility maximization* can be expressed as:

$$U_{ik} > U_{ij} \quad \forall k \neq j, j = 1, \ldots, n \qquad \text{Eq xxviii}$$

where U_{ik} is the utility of a choice-maker *i* making choice *k*; U_{ij} is the utility of the same choice-maker *i* making choice *j*; and $\forall k \neq j, j = 1, \ldots, n$ asserts that *j* stands for all choices other than *k*. Simply then, the above formula establishes a framework for a selection to be chosen from a set of alternatives.

Introducing the idea of *probabilistic* choice-making develops the random utility formula further:

$$P_{ik} = \Pr(U_{ik} > U_{ij}) \ \forall k \neq j, j = 1, \ldots, n \qquad \text{Eq xxix}$$

where P_{ik} is the probability of a choice-maker i choosing alternative k; (Pr is a probabilistic expression). This assigns likelihood to various choices from a set of alternatives.

Adding the assumption that utility may be distilled to a *'strict utility'* and a *stochastic* component yields the final random utility model formula:

$$P_{ik} = \Pr[(V_{ik} + E_{ik}) > (V_{ij} + E_{ij})] \quad \forall k \neq j, j = 1, \ldots, n \qquad \text{Eq xxx}$$

where V_{ik} and V_{ij} are the 'strict utility' components of an individual i's choices of k and j respectively and E_{ik} and E_{ij} are the stochastic elements of the utility calculation for choices k and j. Additional elements may be added to this formula to weight the probability calculation, e.g., variables representing the socio-economic characteristics of a choice-maker.

The random utility model has many similarities to the entropy-maximizing gravity model. There are important differences though. Their similarities may be in large part a function of the set of assumptions upon which they are formulated, rather than their theoretical justifications or actual mechanics. There is also a difference in the way the two approaches handle the assumptions under which they operate, particularly in how they order them. Entropy models assume choices to be random from the outset, then narrow the choice set by applying constraints. Random utility models, on the other hand, start out by assuming a rational choice base and introduce a random element as they proceed (Government of Ireland 1995).

Non-hierarchical logit models

The multinomial logit model is one of the most common techniques for expressing random utility. The logit model expresses the decision choice as a function of the utility of choosing one alternative over another. The model is derived by making assumptions regarding the random component of utility, E_{ij}.

It is commonly assumed that the distribution of E_{ij} follows a Weibull distribution (also known as a double exponential or extreme value type I distribution). The assumption of a Weibull distribution affords the utility calculation a greater degree of mathematical tractability. Applying the Weibull function to the random component of utility leads to McFadden's logit model (Louviere *et al.* 2000; McFadden 1974) in the form:

$$P_{ik} = \frac{\exp[V_{ik}(X_k, S_i)]}{\sum_{j}^{n} \exp[V_{ij}(X_j, S_i)]} \qquad \text{Eq xxxi}$$

where X_k and X_j are the choice-specific attributes of choices k and j respectively (e.g., in terms of trip-making, this could represent costs, time, etc.); and S_i is the individual-specific attributes of choice k (e.g., the choice-maker's income, level of auto ownership, etc.). In short then, the McFadden logit model asserts that the probability of a choice-maker choosing an alternative k from a set of available alternatives is a function of the attributes of the available alternatives and the choice-maker's own characteristics (Government of Ireland 1995).

The non-hierarchical logit formulation suffers from some weaknesses however. Behaviorally, the logit framework assumes that individuals evaluate every available alternative to their choice before settling on an optimal one. In practice, cities generally offer far too many competing alternatives to any given choice to be completely evaluated in this manner. Rather, choice-makers, be they individuals or groups, are more likely to settle on a final choice from a small subset of the available alternatives that are globally available to them throughout the entire urban system. A *hierarchical* choice-making strategy is thus more likely to be employed than an optimizing strategy (Fotheringham & O'Kelly 1989). Structurally, logit models exhibit weaknesses owing to the independence from irrelevant alternatives problem and the assumption of regularity. The problem of independence from irrelevant alternatives (popularly known in transport modeling as the 'red-bus-blue-bus conundrum') stems from the fact that logit models assume that the ratio of probabilities of an individual selecting two alternatives is irrelevant from the addition of an extra alternative. Yet, the introduction of additional alternatives is generally quite relevant in spatial terms. The idea of regularity is closely related to this: the notion that it is not possible to change the probability of selecting an alternative within the logit framework by adding another alternative to the choice set (Fotheringham & O'Kelly 1989). In the context of an urban system this assumption holds little value; it leaves the choice-making process unaffected by any offer of additional choices to a choice-maker.

Nested logit models

The nested logit model departs from the basic logit formulation by introducing hierarchy into the choice-making process. Nested models assume that choice-makers process information about choices in a chained fashion and that the modeler is aware of the form of that chain. In this sense they attempt to circumvent the weaknesses of the non-hierarchical model by assuming that choice-makers make choices sequentially, rather than wading through every available option at once.

Travel choice is a common application of the nested model. Various sequential stages in the decision to travel can be identified, e.g., whether or not to make a trip; where to go; by what mode a trip should be made; and on what route to travel. This method of simulation abstracts from irrelevant (or less relevant) information regarding a choice, and focuses choice on the set of alternatives that are most applicable. Mathematically, this results in a set of conditional probabilities for each of the sequential stages in the choice-making process. Aggregating these probabilities yields the overall likelihood of a choice being made such that:

$$\Pr(a * b * c * d) = \Pr(a) * \Pr\langle b \mid a \rangle * \Pr\langle c \mid a, b \rangle * \Pr\langle d \mid a, b, c \rangle \qquad \text{Eq xxxii}$$

where a, b, c, and d refer to the four sequential stages in the choice hierarchy (e.g., whether to travel, destination, mode, and route). It is important that the estimation process begin with the last step in the hierarchy and work its way back to the beginning in order to ensure that the strict utilities are preserved throughout the process (Government of Ireland 1995).

By formulating the nested model in *spatial* terms, choice-makers in the model are characterized as choosing options from a set of clusters. Continuing with our trip-making analogy, we now have trip-makers (more often groups of trip-makers) who make decisions about their trips but also have to consider a range of spatial options in which to focus those choices. Mathematically, this can be represented as:

$$\Pr_{is} = \frac{\exp(V_{is})\left[\sum_{k\in s}^{n}\exp(V_{ik})\right]^{\sigma}}{\sum_{s}^{n}\exp(V_{is})\left[\sum_{k\in s}^{n}\exp(V_{ik})\right]^{\sigma}}$$ Eq xxxiii

where $\Pr_{is}$ is the probability that a choice-maker i will select a particular spatial cluster s to focus her decision in; $\sum_{k\in s}^{n}\exp(V_{ik})$ is termed an 'inclusive value' and describes the attractiveness of a cluster as a function of the individual alternatives available within that cluster (Fotheringham & O'Kelly 1989); and σ represents the extent to which choice-makers process their information hierarchically, with value between zero and one, and with $\sigma = 1$ denoting choice-makers who do not process their information hierarchically at all.

Once a choice-maker has selected a given spatial cluster, *s*, to narrow her choice set, all that remains is for an option (or alternative), *k*, to be settled upon. The likelihood of a choice-maker selecting a particular alternative *k*, *within the selected spatial cluster s*, is then calculated as:

$$\Pr_{ik\in s} = \frac{\exp(V_{ik})}{\sum_{k\in s}^{n}\exp(V_{ik})} \quad \forall k \in s$$ Eq xxxiv

and the probability of a choice-maker selecting *k from the set of all alternatives* is:

$$\Pr_{ik} = P_{is} * P_{ik\in s}$$ Eq xxxv

Spatial choice models are perhaps more appropriate representations of how urban systems organize than spatial interaction models, but they suffer from weaknesses. Notably, it is widely understood that the distinctions between choice categories may often be fuzzy rather than discrete. Spatial choice models do not commonly accommodate this.

Micro-simulation

Micro-simulation was developed primarily as a tool in public policy, pioneered by Orcott and colleagues (Orcutt 1957; Orcutt *et al.* 1976; Orcutt *et al.* 1961). However, it has been used in urban modeling also (see Clarke 1996 for examples), largely as a response to the generality of many urban modeling methodologies—the division of urban activity into just a few classifications, as, for example, in the case of the Lowry model (Lowry 1964) and input-output analysis from regional science (see Isard *et al.* 1998 for examples). Michael Wegener's large-scale urban model of Dortmund in Germany (Wegener 1983, 1994, 1996) is a well-known example of urban micro-simulation.

Essentially, micro-simulation is a method for disaggregate modeling—disaggregate in terms of entities represented and their attributes, but not always spatially disaggregate. Micro-simulation modeling techniques focus on the probability of events happening to 'decision units', based on a disaggregated set of attributes of that unit. Decision units are themselves often disaggregated in the models; households and firms may be represented, for example, rather than industrial sectors or socioeconomic groups. A variety of events might be micro-simulated. Demographic events include things like births, deaths, immigrations. Family events might be represented as marriages, divorces, changes in household size. Inter-regional migration events are often represented, including changes in the distribution of individuals and families over regions, for example (Isard 1998, p. 403).

Events can also be linked across longitudinal time series. Future attributes of a population, for example, might be determined from a series of successive micro-simulations, each based on results from a prior micro-simulation (Isard 1998, p. 403).

Micro-simulation offers advantages over other modeling approaches, particularly in terms of its attention to classification detail in describing decision units. Nevertheless, the common absence of *spatial* disaggregation is problematic. Also, there is a strong tendency to treat decision units as *average individuals*, with attributes derived from groups and generalized to disaggregate conditions; this invokes ecological fallacy (Openshaw 1983).

Criticisms of traditional approaches

This section examines the methodologies discussed previously, looking at their applicability for modeling urban systems. We can identify several key weaknesses of 'traditional' spatial models in this context, particularly when contrasted with newer models currently being developed and applied: a poor treatment of dynamics, weak attention to detail, shortcomings in usability, reduced flexibility, and a lack of realism. These criteria are used again in Chapter 3, Chapter 4, and to a certain extent in Chapter 5, when discussing 'new wave' models more explicitly. These criteria are used in Chapter 3 and Chapter 4 to assess 'new wave' models.

Dynamics

Spatial models should be capable of capturing the ability of a system or phenomenon to evolve over time. Some important work has been done, extending urban models to take account of dynamics, in terms of rates of change in a 'Systems Analysis' context (Forrester's (1969) 'Systems Dynamics Modeling' of urban systems, for example) and non-linear relationships between elements of urban systems (Allen 1997; Allen & Sanglier 1979). However, for the most part, dynamics have traditionally been relatively poorly represented in the methodologies discussed in this chapter. Dynamics usually enter models in an indirect and implied sense. *Cross-sectional* data are commonly used as a proxy for dynamics. These data are collected for a single period in time: a snapshot. Clearly, this is a poor substitute, but is often the only available option. Other models are developed with *longitudinal* data, offering a series of snapshots, often separated by long periods of time with little information about the intervening period, e.g., data from the Census, which is reported on a ten-year basis. While longitudinal data are much richer in the information they convey, they still constitute a weak proxy for dynamics—for example, a lot can happen in a city in ten years!

Ideally, dynamics would feature more explicitly in a simulation, with system dynamics evolving in real or near-real time (see Gleick 2000 for interesting discussion about near-real-time). Some of the techniques that we will discuss later incorporate dynamics in a more realistic manner—through system-specific time-horizons and interactive dynamics—and offer significant advantages over 'traditional' techniques. This will become particularly evident in the analysis of results of model experiments in Chapter 8, where dynamics feature as one of the key explanatory factors in considering the development of simulated urban systems.

Detail

'Traditional' spatial models are often relatively weak in handling detail. Even micro-simulation approaches, although disaggregate, often deal with 'mean individuals'. More often than not, *spatial* detail is not given careful attention. For the most part, this is due to a lack of data available at fine-scale resolutions and is also a function of their attention to explaining *general* conditions.

In an urban context, 'traditional' models generally adopt the Traffic Analysis Zone (TAZ) as a minimum level of spatial resolution. TAZs are quite aggregate levels of geography: a medium-size city would be divided into just a few hundred TAZs, for example (Figure 2). From this level of geography, one can only *infer* information at the level of individuals or entity-level geographies of urban space and to do so invokes issues of ecological fallacy and modifiable areal unit problems (Openshaw 1983). Because many of the processes that 'make cities work' operate at finer resolutions, this lack of detail may, in some cases, be regarded as a serious limitation of 'traditional' spatial models for urban applications.

Figure 2. Spatial resolution: Megalopolis and New York

This is difficult when there are not adequate data to support the required level of detail. Nevertheless, as will be discussed in more detail in Chapter 3, more detailed data-sets are becoming available for use and over time they will be accessible for historical periods, enabling the calibration of fine-scale micro-simulations. ‘Traditional’ techniques often lose efficiency as the level of detail increases, specifically as the matrix of relational entities in the model grows. In Chapter 4 we will explore a series of techniques that embrace detail in a more integrated fashion than ‘traditional’ techniques and offer the potential for a more resourceful handling of detailed data.

Usability

Many urban models are used in practice to inform decisions (Brail & Klosterman 2001; Geertman & Stillwell 2002). Usability is therefore another issue that needs to be considered when assessing the value of a simulation methodology for urban application. Usability has long been a concern in other areas of applied science, e.g., human-computer interaction in computing (Preece 1994); but has often been weakly addressed in operational urban simulation. In many cases, users perceive simulations as ‘black boxes’: inputs are fed into the model and results are output, but the inner workings of the model may remain a mystery. This acts as a barrier to the efficient and appropriate use of models as decision support systems and impairs the ability of models to serve as exploratory tools. The techniques introduced in Chapter 3, Chapter 4, and Chapter 5 are perhaps more user-friendly than those mentioned in this chapter.

Spatial flexibility

Often, it is important that urban models accommodate a wide variety of spatial scales, ideally in an integrated and seamless manner that is capable of representing the phenomena that shape urban areas at all levels from global through to local scales. As we have seen, 'traditional' spatial models are weak in their handling of micro-scale phenomena. Some of the methodologies introduced in Chapter 4 and Chapter 5, supported by advances mentioned in Chapter 3, can better facilitate modularization and lend models greater flexibility. The expression of models in object-oriented terms, in particular, offers the potential for an improved level of flexibility. This approach is discussed, explicitly, in the formulation of the Geographic Automata Systems framework detailed in Chapter 5.

Realism

Bluntly stated, cities do not always work the way that 'traditional' spatial models would have us believe they do. As discussed throughout this chapter, there is often a disparity between models and reality on a behavioral level. In particular, 'traditional' spatial models adopt a reductionist view of systems. For the most part, assumptions are made that portray cities as operating from the top down. Even the conceptual structure of 'traditional' spatial models betrays a bias in their formulation: models are often illustrated as flow diagrams that begin with a regional scale model and filter down to TAZ-level components. With the exception of a few feedback mechanisms, all of the arrows point downward. The reductionist approach implies that to understand urban systems, it is necessary to dissect them into constituent local components from aggregate conditions. In many cases, this is perfectly right! However, in other instances it is inappropriate. Many components of urban systems (planning and public policy, for example) do not work in a top-down manner; on the contrary, aggregate conditions emerge, from the bottom-up, from the interaction of large numbers of elements at a local scale (Holland 1998). In the cases of bottom-up system dynamics, 'traditional' spatial models often run in the wrong direction. The techniques mentioned in Chapter 4 proceed from the bottom-up. The examples developed in Chapter 8 and Chapter 9, however, adopt a more bi-directional approach.

Conclusions

This chapter has presented a wide-ranging overview of what we might regard as being 'traditional' spatial modeling methodologies. The discussion has examined popular techniques for modeling various spatial properties of urban systems. The field is large, and the overview is understandably general. Particular attention was paid to *critiquing* traditional methodologies. This paves the way for the introduction of a 'new wave' of simulation in the remaining chapters, catalyzed largely by recent developments in computing sciences and the geographic sciences. Nevertheless, the techniques discussed in this chapter remain the cornerstone of urban simulation. This chapter has focused on critiquing those methodologies, but as later chapters will illustrate, the techniques filter into much of the 'new wave' methodologies that will be discussed.

The review in this chapter has been quite critical of traditional modeling approaches. It is important to recognize the strengths of that stream of work, however. First, and perhaps foremost, traditional models have been around for some time and have enjoyed popular use—in a wide variety of disciplines. By contrast, the techniques described in chapters 3, 4, and 5 are at a largely experimental stage in their development—in many instances they have not been applied in practice to the same extent as more traditional methods and the theory that underpins more conventional approaches has not been as fully developed in the context of urban systems. This is particularly evident with respect to validation schemes—validation is generally well-developed for traditional approaches; that is not the case to the same extent for newer approaches.

Second, it is important to consider the critique in the context of what the models are actually designed to do. Traditional approaches were, to a large degree, developed to describe urban systems from a largely static standpoint and at an aggregate scale of consideration; they do this job admirably. Additionally, there is a stream of research represented in the literature that has extended that approach to in corporate dynamics (see Allen (1997) for example) and development in this area has been significant. Indeed, much of that work has set the stage for the introduction of automata-based techniques into the field, through emphasis on treatment of urban systems as complex phenomena.

Third, it is worth noting that traditional techniques are quite compatible with newer approaches, although work in this area is just beginning. Newer approaches such as automata techniques operate well at lower-scales of consideration, but need some tight constraints to generate sensible results at meso- and macro-scales. This is simply a function of the degree to which the micro-geography of urban systems is largely unknown in many cases. There is much room for connecting more traditional approaches that handle aggregate conditions and operate from the top down with particular grace, with local-scale and bottom-up tending techniques based on automata ideas. This is actually what is done, on a mostly conceptual level, in the models described in chapters 8 and 9, where growth rates, hypothesized to come from a higher-level, essentially set the micro-models into action.

The difference between traditional techniques, as described in this chapter, and 'new wave' approaches is discussed in Chapter 4. The models developed in later chapters of the thesis have been designed with careful consideration of the issues raised here. The Geographic Automata Systems framework that is developed in Chapter 5 was designed with careful consideration of the topics discussed in this chapter. The sprawl model built in Chapter 8 is concerned, largely, with overcoming the limited theoretical justification of spatial interaction models, particularly at micro-scales, where attention turns to individual trips and movements—within a flow—in urban space, and the actual *choreography* of that movement, from a behavioral basis. Decision and choice form the basis for describing the behavior of simulated entities in the model described in Chapter 9; again at a micro-scale and with consideration of spatial and dynamic aspects of the system.

Chapter 3. Conditions supporting a 'new wave' of urban simulation

"The universe was a disorderly mess, the only interesting bits being the organized anomalies." (Stephenson 1995, p. 63)

Introduction

Chapter 2 detailed what we might term, 'traditional' spatial methodologies. Those methodologies were discussed critically, evaluating spatial simulation techniques in terms of their suitability as laboratories for exploring ideas about urban systems.

The discussion in this chapter is focused on the conditions supporting a 'new wave' of urban simulation that has emerged in recent years, very much as a reaction to the kinds of issues that were raised in Chapter 2. Specific methodologies and their use in simulating cities are discussed in detail in Chapter 4. The purpose of this chapter is to explain the forces behind the development of new methodologies for simulating urban systems, to demonstrate the threads of research in the field devoted to remedying the issues mentioned in Chapter 2, and to serve as a foundation for the introduction of automata-based tools in Chapter 4 and the remainder of the thesis thereafter.

This chapter focuses on the developments in geography and outside the field that have catalyzed the infusion of powerful new simulation techniques in urban contexts. The discussion continues in the next section with mention of dataware—information and information processing resources—for urban modeling. The discussion focuses on the contribution of GI Science, spatial analysis, GIS, and remote sensing to new simulation approaches, and mentions very recent trends in dataware, with particular treatment of synthetic data generation routines. Following that, the contribution of complexity studies to 'new wave' simulation is considered, both in terms of systems thinking and its impact on the conceptualization of urban systems, and particular methodological contributions, including those from Artificial Life. Next, software resources are discussed, with particular attention to the influence of object-oriented programming and software modularity on simulation development in urban contexts. That discussion has particularly strong implications for the Geographic Automata Systems discussed in Chapter 5. Recent trends in the development of general software resources for fine-scale dynamic simulation in the social sciences are also discussed. The discussion then shifts to an evaluation of the relative impact of these developments on urban simulation, evaluating the potential of recent advances to extend the functionality of urban modeling with respect to spatial scale, time and dynamics, and the representation of systems. Much of this discussion extends threads begun when critiquing traditional approaches in Chapter 2. Taken together, the developments discussed in this chapter are considered in terms of what might be regarded as a burgeoning paradigm shift in urban simulation.

Dataware

Advances in dataware—information and data collection resources; tools for querying, manipulating, analyzing, and visualizing data; methodologies for generating data—have been particularly influential in advancing the potential for urban simulation. Advances in dataware have supported a new wave of urban simulation by providing techniques and tools for analyzing and monitoring urban systems in new ways, and with increased spatial and temporal resolutions. They also allow urban models to be fed with new sources of information about urban systems. Progress in the geographical sciences (GI Science, GIS, and spatial analysis) and geomatics (remote sensing) have been particularly influential, as have other developments, such as the derivation of synthetic populations.

GI Science, spatial analysis, and GIS

GI Science, GIS, and spatial analysis have provided a range of methodologies for handling, interpreting, and producing data for urban simulations. Traditionally, they have been used to analyze urban data and to provide information—variables—for urban models, or have been used as calibration and verification tools to register simulation runs against real world conditions. These are loose-couplings, indirect connections. Data may be generated in a GIS, for example, and then fed into an urban simulation. The output that is generated by the simulation could then be fed back into the GIS to be visualized, or to have spatial analysis performed on it.

There also exist opportunities for tight-coupling GI Science, spatial analysis, and GIS with urban models. Models can be designed in such a way that GIS form the database architecture for a simulation. Simulation runs could, for example, be called from within GIS. There is also potential for incorporating GI Science and spatial analysis directly into urban models. GI Science and spatial analysis provide formal methodologies for expressing relationships between geographic objects, as well as representing the behavior of objects and processes in space. (This actually forms the basis for several of the components of Geographic Automata Systems discussed in Chapter 5.) All of these functionalities support the exploration of geographic patterns, trends, and relationships in urban systems.

Remote sensing

Developments in photogrammetric engineering have provided an assortment of tools for collecting data about cities from remote platforms: airborne and satellite sensors. These advances have provided new information about urban systems, and in many instances remotely-sensed data can be fed directly into urban simulation models.

The resolution and coverage available through remote sensing is growing steadily, offering new insights into urban systems and providing fine-scale datasets for urban models over a wider spatial range. Moreover, these technologies have been in place for a long time, and now offer longitudinal data across long periods of time.

In tandem, innovations in image processing have facilitated the derivation of information from remotely-sensed data. Development of methodologies for automated feature-extraction has been particularly useful. Land cover and land-use classification schemes enable the inference of socioeconomic information from remotely-sensed images. Likewise, a range of techniques exist for elucidating urban morphology, from the identification of individual structures to the interpretation of digital signatures for various configurations of urban infrastructure (see Longley & Mesev 2001; Torrens 2004b; Webster 1995).

Synthetic data generation

Research into the generation of synthetic data sources is also beginning to emerge. The generation of synthetic population for the application of the TRANSIMS model (Barrett et al. 1999) to Portland, OR (Barrett et al. 2001) is one example. In that example, a range of methodologies was used to generate realistic and statistically-fit population data, at a micro-scale. Synthetic households were generated, with demographic and socioeconomic attributes, geo-referenced to micro-scale geographies. The TRANSIMS Population Synthesizer Module (Barrett et al. 2002) and the Geographic Correspondence Engine (Blodgett 1998) were used to generate the data. They took a variety of readily available inputs: Census data, publically-available micro-samples (the Public Use Micro-sample), TIGER polygon boundaries, tax lot data, block-level demographic forecasts, and vehicle records (Bush 2001).

Complexity studies

Complexity studies are perhaps one of the most important developments outside geography and urban studies to have influenced urban simulation in recent years. Complexity studies, as a field, offer a new theoretical framework for thinking about how urban systems function and how they might be simulated (see Allen 1997). Much of the theory associated with complex systems has begun to filter into urban modeling and the methodology associated with it. Complexity studies have influenced the way we think about—and model—systems; there is a move toward representing systems as networks of inter-dependent components. This is not new; the idea was popularized by Forrester many years ago (Forrester 1969). However, complexity studies have also led many model-builders to think about urban systems as being *interactive* and *adaptive*, and models have begun to simulate cities as organizing from the *bottom-up*.

The idea of complexity hinges on the notion of emergence. Complexity studies have been termed as the *science of emergence* (Krugman 1996). In emergent systems, a small number of rules or laws, applied at a local level and among many interacting objects or agents, are often capable of generating surprising complexity in aggregate form. The patterns that complex systems generate often manifest themselves in such a way that the actions of the parts do not simply sum to the activity of the whole (Holland 1998). Emergent phenomena share some key characteristics, even when compared across disciplinary boundaries. They are commonly dynamic and subject to change over time. Emergent properties evolve through the interactive dynamics of the system. Emergent systems often function without the direction of a centralized executive; the systems can sometimes be regarded as *self*-organizing (Allen 1997; Portugali 2000). Also, emergent systems often generate ordered features.

Examples of emergent systems abound. Stock markets are a good example: markets such as the New York Stock Exchange (NYSE) are comprised of millions of traders buying and selling in a bid to maximize their own individual profits. In the eighteenth century, the Scottish economist Adam Smith postulated the idea of an "invisible hand" that set the level of equilibrium between supply and demand in the market place (see Krugman 1996). This hand was just a metaphor, but a striking one: individual investors in stock markets act without any centralized control, yet their activities often lead to aggregate outcomes that are relatively efficient, as efficient as if they *were* controlled.

The main value of the complexity approach for urban simulation is its emphasis on detailed, non-linear, and bottom-up approaches to understanding urban systems. This is a relatively new way of approaching scientific inquiry in social science. Much research in the social sciences, and particularly in geography, is challenged by a dichotomy between the individual (the household, a person, and independent objects) and the aggregate (populations, collectives, and regions). In a spatial sense, researchers have been confronted with the dilemma of reconciling patterns and processes that operate and manifest at *local* scales with those at *larger* scales. This spawns problems of ecological fallacy (Wrigley et al. 1996) and modifiable areal units (Openshaw 1983). There are many examples in which aggregate forms may be extrapolated from the individual. However, reconciling the two often poses a challenge, particularly when processes that operate at the local level are interdependent, i.e., the actions of one individual depend on the actions of another individual. In these cases, an understanding of the processes that generate macro-scale patterns may not be easily gleaned by simply aggregating up from the individual; what is needed instead is an understanding of the *interactive* dynamics that link local-scale and larger-scale phenomena.

Reductionist approaches analyze problems by breaking them down to their constituent components, reducing them to manageable pieces and gaining an understanding of them in the process. In some cases this approach works quite well, and for many phenomena the technique is wholly appropriate, particularly in situations where the whole is the sum of many smaller parts. However, the reductionist approach is flawed in the respect that it may miss the emergent properties of a system: those that come as a by-product of the *interactive* dynamics of individual elements. In many instances, a generative method may be more appropriate. Generative approaches involve studying phenomena by experimenting with simple rules for behavior and allowing constituent components to interact, dynamically, until macro-scale phenomena emerge—a piecing together rather than a dissection (Taylor 1992). This is what happens in our own bodies. The rules encoded in our DNA specify a set of behaviors for the development of our biology over time. The products of that interactive development on a genetic level are apparent at a macro-level as distinct structures—organs, systems, and traits—that bear little resemblance to the original components of our DNA. The central nervous system, for example, is significantly more complicated than the arrangement of bits of guanine, adenine, thymine, and cytosine along a genome. Researchers are increasingly adopting generative approaches to the study of phenomena, particularly in studying life, where it has been noted that,

> "Reductionism does not work with complex systems, and it is now clear that a purely reductionist approach cannot be applied when studying life: in living systems the whole is more than the sum of its parts." (Levy 1992, p.8)

There are many reasons why we might transfer ideas from complexity to our understanding and conceptualization of cities. From the local-scale interactive behavior (commuting, moving) of many individual objects in cities (vehicles, people), structured and ordered patterns emerge in the aggregate, such as peak-hour traffic congestion (Nagel et al. 1996) and the large-scale spatial clustering of socioeconomic groups by residence (Schelling 1978). In urban economics, large-scale economies of agglomeration and disagglomeration have long been understood to operate from local-scale interactive dynamics (Krugman 1996). Also, cities exhibit several of the signature characteristics of complexity, including fractal dimensionality and self-similarity across scales, self-organization, and emergence (Allen 1997; Batty & Longley 1994; Portugali 2000).

Artificial Life

As a discipline, Artificial Life (A-life) is concerned with the possibilities and potential for creating synthetic life-forms. A-life studies are directed toward understanding natural life and attempting to abstract the fundamental principles of life; a search for the rules that make life possible. For the most part, applications of A-life focus on creating synthetic life *in silico* (Levy 1992), in computer simulations. This is largely motivated by desire to replicate (or generate) the complex dynamics of living organisms and systems in other media. In this sense, it is hoped that the principles that govern real life can be uncovered in the process.

A-life studies have several parallels with and connections to urban simulation and A-life has contributed greatly to urban simulation. A-life offers a new conceptual framework for urban models. The focus in much A-life research is on bottom-up dynamics—emergence. A-life emphasizes the interaction between divergent entities and systems. This is, of course, the way we understand urban systems to function. Indeed, cities are popularly referred to as organisms (Jacobs 1961). Much of the conceptual methodology from A-life research has influenced urban modeling, particularly in relation to inter-system dependencies and human-environment interaction.

In some instances, simulation methodologies have been borrowed directly from A-life and have featured in urban models. Boids (Reynolds 1987) and Animat (Meyer et al. 2000) movement and behavior routines have been ported to traffic models, as have swarming (Bonabeau et al. 1999) and flocking mechanisms (Torrens 2004a). The strong emphasis on computational experimentation has, perhaps, had some influence in advertising the utility of simulation as a planning support tool (Torrens 2002). Several A-life toolkits have also been adopted by urban modelers and adapted to urban uses; cellular automata and multi-agent systems are a good example of this, and we will discuss them in more detail in the next chapter.

Finally, the interdisciplinary nature of A-life research has had an indirect influence on urban simulation. A-life work spans several disciplines: the computing, physical, life, behavioral, and social sciences. Urban modelers working in A-life research draw upon all of these fields in their work, connecting to fields of study outside their traditional home discipline.

Software

Urban simulation has also benefited greatly from advances in computer software: new paradigms for writing simulation software, developments in the inter-operability of diverse software components, and the release of a diversity of resources for analyzing and modeling cities.

Object-oriented programming

Object-oriented (OO) programming is a method of writing software in which basic units are combined to accomplish tasks. OO programming languages include Smalltalk, Java, C++, and C#. Programs are designed in such a way that their code is represented as collections (classes) of objects. Objects are simple containers, for data, variables, other pieces of code, etc. Objects are defined by information and methods (behaviors) that are associated with them. Methods are usually formulated as conditional statements that determine how objects should interact and evolve over the course of a program run (see Horstmann & Cornell 2002; Hortsmann & Cornell 2001 for more details).

The conceptualization of pieces of inanimate model code as objects with related data and methods mimics the way that we think of real world objects ourselves: as discrete units with associated attributes and behaviors. This has several benefits for making models more flexible to build, as well as making them easier to convey and understand. OO approaches offer obvious advantages for the treatment of discrete entities in urban systems. The objects that comprise urban systems, such as land parcels, buildings, administrative zones, households, and individuals, can be simulated in the program in ways analogous to the way in which they appear in the real world, and conditional statements can be designed to mimic how they behave and interact.

Software modularity

Urban modeling has also been influenced by developments in software modularization. A variety of tools is now available for coupling diverse software resources and programs. This enables models from a variety of disciplines to be designed to simulate a range of urban systems (or even non-urban systems), and to be connected in a symbiotic manner.

The Component Object Model (COM) is one of the most popular modularization tools (see Microsoft Corporation & Digital Equipment Corporation 1995). COM facilitates the coupling of separate pieces of software, e.g., two models. Model A, for example, can be run, generating data. That data may then be passed through COM, which invokes Model B to run further simulation using the data. The results may then be fed back to Model A. Other modularization schemes, e.g., Java Beans (Horstmann & Cornell 2002), work in a similar manner. With Beans, your code is sheathed in a 'wrapper' that allows it to interoperate with other, similarly-wrapped code.

Modularization in this manner has several practical uses for urban simulation. Urban models can be connected to other urban models, e.g., a land-use model may be fused with a transport model, or various modules devoted to individual urban subsystems may be bound together in a common architecture (Noth et al. 2003). Urban models can also be connected with non-urban models, e.g., socioeconomic and environmental models (Alberti & Waddell 2000). Urban simulations can also be connected to other software packages, e.g., GIS, spatial analysis (Ungerer & Goodchild 2002), statistics, or visualization packages.

Available software resources

Independent software packages for urban simulation are generally not widely available, largely because urban modeling is a nîche market and, for the most part, it is an application-specific enterprise. However, there is an abundance of "off-the-shelf" software products to *support* urban simulation. Moreover, the popularity of the open source movement (Moody 2001) is extending to urban simulation. Several open source dataware packages are now available, and a handful of open source modeling tools—and models—are beginning to enjoy widespread use (for examples, see http://www.digitalearth.org).

A number of freely-distributed dataware products are being popularly used, e.g., GRASS (a GIS package) (Baylor University 2002) and FRAGSTATS (a landscape ecology and spatial analysis package) (McGarigal & Marks 1995). Open source libraries such as GeoTools (a lightweight GIS toolkit) (Centre for Computational

Geography 2003) are beginning to be coupled, modularly, to models, opening up a range of possibilities for integrated dataware-simulation development. Several open source code libraries for general simulation are also beginning to be used by urban modelers; Repast (University of Chicago 2003), Ascape (Brookings Institution 2001), and Swarm (Swarm Development Group 2001)—all agent-based toolkits for social science simulation—are particularly popular. In addition, several macro-language simulation products are freely available and are beginning to be used to build urban simulations, e.g., StarLogo (Resnick 1997), StarLogoT (Center for Connected Learning and Computer-Based Modeling 2001), and NetLogo {(Center for Connected Learning and Computer-Based Modeling 2003b). (Both StarLogoT and FRAGSTATS are used to develop and validate the models described in part three of the thesis.)

A 'new wave' of urban models

Together, these developments have catalyzed a 'new wave' of urban models (Benenson & Torrens 2004a). The traditional toolkit of models is rapidly being supplemented—and in some instances replaced—by a next generation. Whereas the traditional toolkit might be characterized as relatively coarse, static, and inflexible, the trademarks of a 'new wave' are largely antithetical to this characterization. Recent innovations in urban modeling have facilitated significant advances in the spatial and temporal scales of city simulations, the representation of spatial processes and spatial behavior, and the treatment of complex systems. The technology of urban simulation has also undergone a transformation in terms of concurrency, transparency, and inter-operability. Finally, these developments have led to a burgeoning paradigm shift in urban simulation, greatly expanding the usefulness and range of uses of urban models.

The upcoming discussion will focus on the merits of new wave approaches. However, it is worth noting that, in many ways, more traditional approaches can complement newer-style techniques. This is particularly true at macro-scales and for systems that operate from the top-down; areas where more traditional approaches generally shine. Newer techniques are excellent in their own realm—the micro-scale and massively interactive arenas that function with bottom-up tendencies. However, to work at larger scales or to accommodate top-down dynamics, urban models of this nature more commonly turn to methods or inputs of a more traditional nature.

Spatial scale

The aforementioned advances have had a significant influence on urban modeling in terms of spatial scale. Conventionally, urban models have been designed to operate at relatively coarse spatial resolutions. As discussed in Chapter 2, spatial units are often represented at broad geographies; Traffic Analysis Zones (TAZs) are common. This reliance on coarse units often limits the usefulness of urban models as exploratory tools; in some instances, the results that are generated may be too general to be useful. The infusion of new dataware and theory has supported the development of urban models with much finer resolutions. Simulations can now be run, representing urban systems at the scale of individual entities. Individual pedestrians, vehicles, households, buildings, and land parcels now replace TAZs as the base of urban simulations—atomic units (in the sense that they deal with individual entities such as people and cars) rather than artificial aggregations. This enables the specification of much more detailed urban models, but it also facilitates the representation of individual-scale behaviors and interactions. Moreover, should aggregate outcomes be required, they can be generated, sensibly, from the bottom-up. The models developed in part three are indicative of this.

Increases in spatial resolution represent a significant leap in the realism of urban models. Individual units can be represented, but so too can their individual and *independent* characteristics and behaviors. The notion of the 'mean individual', with average attributes derived from the group (Anderson 2002) is steadily being abandoned. This has important implications for circumnavigating problems of ecological fallacy in model development, because the models can be run with spatially non-modifiable units. Several of the models discussed in Chapter 7 operate at fine spatial resolutions; the model specified in Chapter 9 operates at the scale of individual households.

Model developers can 'zoom into' a TAZ. They can also lay out the mechanisms that explain the formation or functionality of an urban system within that zone, focusing on the interactions among individual elements that give rise to the system. It is now possible to move beyond a reliance on interaction as flows between modeled entities—an approach characterized by gravity and spatial interaction models—and into the representation of more *localized* interaction. Because of the increase in the fidelity of 'new wave' models, a variety of individual-scale interactions can be simulated: local movement, migrations, collisions, etc. Higher-level interactions can also be represented, often seamlessly, as they emerge from more micro-scale activity. In addition, this allows model developers to abandon the notion that interactions take place evenly across a system (Anderson 2002).

Time and dynamics

The developments discussed in the preceding sections of this chapter have catalyzed a 'new wave' of simulation through improvements in temporal resolution and dynamics. Traditionally, urban models were often specified as static or comparatively static (simulation runs proceed from one snapshot in time to another, with a large period separating the two points and with little information about the intervening period). By contrast, the latest generation of urban models is highly dynamic, specified at fine temporal resolutions, often near-real-time.

The infusion of simulation methodology from A-life research has been particularly influential in this regard, introducing a variety of techniques for representing fluid-like dynamic interaction and characterizing simulated time as 'packets of change' (Anderson 2002). Complexity studies have also supported the characterization of emerging conditions and dynamic (and non-linear) feedback in urban models. Simulated systems can be designed to react to conditions within the simulation, as they evolve over time. Furthermore, a variety of dataware resources are available to support dynamic urban simulation. Coupled with advances in spatial fidelity, improvements in the representation of dynamics has opened up promising new avenues of inquiry into the space-time dynamics of urban systems, using urban models. This is discussed in more detail with reference to the models developed in Chapter 8.

Systems thinking

The advances already discussed have also impressed upon consideration and representation of systems in urban models. Emphasis on the interdependency and symbiosis of urban systems—across scales—is particularly noteworthy, as is the representation of bottom-up system dynamics, and a growing attention afforded to consideration of evolution of systems and non-linear relationships in simulations. Complexity studies have had a huge influence on the conceptualization and specification of systems and systems components in urban models. In parallel, developments in software have provided the technical mechanisms for developing simulations like this in practice.

The influence of the bottom-up concept on systems thinking in urban simulation has been particularly profound. It has emphasized the need to look at the local interactions within a system: the local links between the entities that comprise a system, and the local space-time dynamics of their connections. The impact of emergence on urban modeling has also been important, catalyzing interest in the ways in which systems, structures, patterns, behaviors, etc. form from independent components and across a diversity of scales.

Finally, the developments discussed in this chapter have led to somewhat of a shift in the way that system dynamics are treated in urban models. There is a strong emphasis, in 'new wave' models, on system evolution, with model developers' and users' developing a growing interest in things like path dependency, the importance of initial conditions, bifurcations, the rate of system evolution, amplifiers, dampeners, etc. Again, those issues will re-surface with discussion of the models reviewed in Chapter 7 and designed in part three.

A paradigm shift in urban simulation

Fundamentally, this has resulted in somewhat of a paradigm shift in urban simulation. Urban simulation has been rejuvenated as a research activity; the inter-disciplinary nature of urban modeling has grown in parallel. Moreover, urban simulation has enjoyed re-branding as an applied exercise and a research methodology.

The developments discussed in this chapter have, through their varied influence on urban simulation, revitalized the field as an area of research. Exploration in the development and application of urban models has become very active after a period of relative dormancy. Some aspects of urban simulation research have become the 'newest new thing' in geography; automata modeling, in particular, is enjoying widespread popularity at the moment. There is a general sense that several lines of research inquiry in the geographical sciences are converging, with the potential for urban simulation to capitalize upon developments in fields like GI Science, GIS, geocomputation, spatial decision support systems, spatial complexity, and geographic visualization. Nevertheless, the research agenda is very much in its infancy.

Inter-disciplinary research is at the core of this apparent paradigm shift in urban simulation research. Urban modeling has always been an inter-disciplinary exercise. However, model-builders are beginning to draw upon a much wider base of knowledge and practice to construct the next generation of urban models. Researchers are connecting with previously-ignored fields of study. Urban simulations have always used *computers*, for example, but are increasingly relying on intellectual resources from *computer science*—AI, A-life, and OO programming, as we saw before, as well as concepts from database theory, animation, human-computer interaction, and parallel computing. Connections with complexity studies have also drawn urban modelers into contact with new fields in the physical and life sciences. The reaffirmation of links with other social sciences have been, perhaps, most significant. Geography is beginning to take center stage in the new field of social science computing, where other social scientists—economists, sociologists, political scientists, anthropologists, archaeologists—are finding that space matters. It is also noteworthy that several of the *quantitative* developments in urban simulation (and the geographical sciences) are enjoying popular use in social science computing.

Fundamentally, the apparent paradigm shift in urban modeling is manifesting itself as a re-branding of simulation as a methodology for experimentation and exploration in geographic research, and in new ways of considering city simulations as applied tools for evaluating plans and policies. 'New wave' urban models are powerful as a tool for exploring theories and ideas. Simulations can serve as an artificial laboratory for testing hypotheses, but with unprecedented degrees of realism and attention to detail. To a certain extent, experimentation with urban models is now limited only be the theories and data that support them; the technology of urban simulation has advanced to the stage where urban models have the potential to simulate most urban systems. (Of course, actually building those simulations is an entirely different undertaking!)

Aspects of urban systems that were previously 'un-modelable' are now within our grasp: real-time dynamics, parallelism on the scale of a city population, etc.

The paradigm shift is also evident in the application of urban models as planning support tools, where there is a shift from considering simulation as a predictive exercise to treating urban models as 'tools to think with' (Torrens 2001, 2002). This is partially a response to advances in urban modeling that have enabled simulations to be developed as intuitive, transparent, and immersive tools. Advances in scale, dynamics, and systems thinking in urban simulation have furthered the *usefulness* of urban models as planning support tools.

Conclusions

This chapter has detailed recent developments in the geographical sciences, and in areas of study beyond geography, that have supported the development of an apparent 'new wave' of simulation in urban contexts. Considered together, these advances represent somewhat of a burgeoning paradigm shift in the field, greatly extending the capabilities and potency of urban simulation as both an exploratory methodology, and a tool for supporting management, planning, and policy decisions in the real world.

Within the geographical sciences, dataware resources have been particularly influential. GI Science, spatial analysis, GIS, remote sensing, and synthetic data generation routines have provided new data sources to 'feed' urban models, as well as offering methodology for extracting detailed and dynamic *information* from those data. Developments in complexity studies have been similarly significant, both in terms of offering new insights into the way we consider and conceptualize urban systems as evolutionary, interactive, dynamic, and complex adaptive systems, and through the infusion of methodology directly from related fields, such as Artificial Life. The computing sciences have been especially significant in supporting recent advances in urban simulation. Indeed, the boundary between geography and the computing sciences is blurring in some areas of the discipline—a new field of *geocomputation* is now established, bridging gaps between the two. The adoption of object-oriented paradigms by geographers has been influential in rejuvenating modeling work, as have developments in software modularity. Furthermore, the publication of several general purpose modeling tools, developed by researchers in the emerging *social science computing* field is beginning to have an impact on the development of urban modeling software.

In the next chapter, we will explore some of the most popular methodologies that have emerged from these developments—automata tools—and we will discuss the potential that they offer for developing new forms of urban simulation, These tools form the basis of a new framework for urban modeling—Geographic Automata Systems—which will be described in Chapter 5, and which serves as the foundation for the development of sprawl models discussed in Chapter 8 and Chapter 9.

Chapter 4. Cellular Automata and Multi-Agent Systems

"*Yes*, I used a nice avatar." (Stephenson 1993, p. 72)

Introduction

Innovations that have supported advances in urban simulation were discussed in the last chapter. The advances that were mentioned have led to the development of new urban modeling tools. Work in the design and application of tools formulated on automata methodologies has been at the vanguard of that field of urban simulation research. Two classes of automata tools—cellular automata and multi-agent systems—have been particularly popular; their use has dominated the research literature with a recent flurry of activity. Both classes of tool stem from the developments discussed in the last chapter and offer significant advantages for general simulation, and urban simulation in particular. However, automata have their origins as computing media, and must be modified significantly for use in urban and geographic contexts. This point is important and, in this chapter, attention is given to the computing foundations of automata before considering their use as simulation—and spatial urban simulation—tools.

This chapter discusses the use of automata tools as a simulation mechanism in urban research. The origins of cellular automata and multi-agent systems as computing media are discussed, in turn. This is followed by an evaluation of their use as simulation tools. In particular, the benefits they offer over the conventional methodologies discussed in Chapter 2 will be highlighted. Following this, the use of cellular automata and multi-agent systems as *urban* simulation tools will be discussed, although specific treatment of their use in modeling suburban sprawl will be reserved for later chapters. Automata form the basis for the Geographic Automata Systems framework outlined in Chapter 5, which is used as the foundation for sprawl models described in part three of the thesis, and developed with consideration of characteristics of the phenomenon, described in part two.

Automata: an overview

Simply stated, automata are independent processing devices. The function of an automaton is to process *information*—internal information that is contained within the automaton itself, as well as information that is input to it from external sources. Automata are capable of processing information independently (Figure 3), but may also be networked to form parallel processing arrays (Figure 4).

Automata have their origins in the specification of the first digital computers. Indeed, all central processing units (CPUs), like those found in personal computers, digital televisions, and electronic calculators, are automata. Fundamentally, automata are computing media, although they have enjoyed a widespread and varied use; notably, in the context of this discussion, they are very useful as simulation tools.

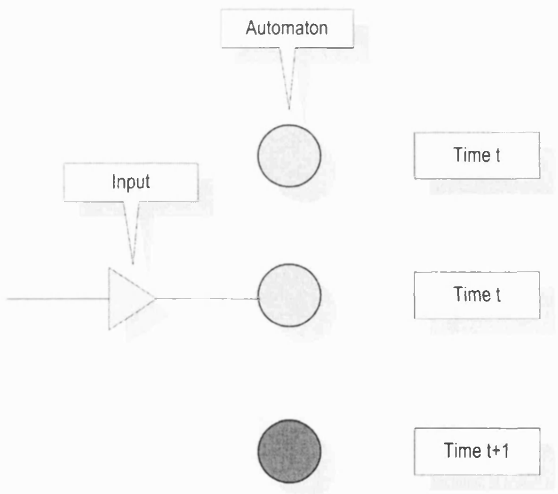


Figure 3. A single general automaton (the colors in circles represent states)

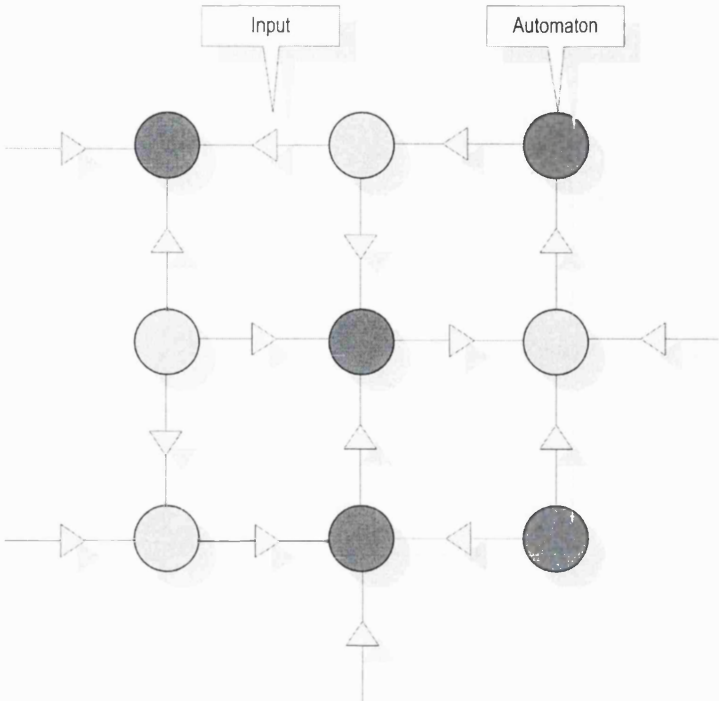


Figure 4. Multiple networked general automata

Automata characteristics

Automata are commonly characterized with the following attributes. Individual automata are housed in some form of *unit*. In the case of tangible, physical automata, the unit generally takes the form of a CPU, such as a computer chip, etched on a silicon wafer. Other physical automata have been housed in organic molecules (Benenson et al. 2001). Non-physical automata, such as those replicated in computer programs, may be housed in a variety of (artificial) structures, such as arrays, polygons, and strings.

The attributes of a given automaton are expressed as a set of *state variables*. States describe the characteristics of an automaton at a particular point in time. States are commonly binary in physical automata—on or off. State variables can be formulated with virtually any data type, however.

Automata *neighborhoods* are localized areas around independent automata, from which they draw input. In networks of automata, neighborhoods are formed as sets of adjacent automata, with which information exchange takes place (Figure 4).

Transition rules govern how automata states alter over time, e.g., whether an automaton should be in a certain state from one transition point to the next. Automata make use of transition rules to evaluate their present states and any information input to them from their neighborhood, thereby determining the future state of the automaton in a subsequent step. Transition rules are best thought of as conditional statements: if 'something', then 'something'; else, 'something'.

For physical automata, such as microprocessors, these rules are expressed as Boolean gates, e.g., NAND, NOR, NOT, AND, OR, XOR. These gates allow for transition rules to be mapped into bits and bytes and computed on a microprocessor. Essentially, they allow for the evaluation of conditional statements and mathematical operators. In non-physical automata, a variety of rules may be specified. In terms of simulation, rules can be designed to mimic the laws or processes that govern events and activities in a phenomenon of interest.

In addition, we can think of *time* as a component of automata. In automata contexts, time moves in discrete steps; transition points serve as individual moments in time, or packets of change. The Earth Simulator, developed by the National Space Development Agency of Japan, the Japan Atomic Energy Research Institute, and the Japan Marine Science and Technology Center, is the world's fastest supercomputer at present. The Earth Simulator is, in fact, a sophisticated automata array—a cluster of networked CPUs. It can perform 35 teraflops—35 trillion (10^{12}) floating point operations per second. Each of these calculations may involve anywhere from hundreds to hundreds of thousands of independent state transitions per calculation.

Automata as computing media

Automata were first conceived of in the 1930s by British mathematician, Alan Turing. Turing *hypothesized* about an automaton machine, which would come to be known as the Universal Turing Machine. The Machine was to be specified with a limited range of attributes. Conceptually, the Machine would be capable of universal computation. Given a suitable initial program (transition rules), such a machine should be able to produce a working copy as complicated as itself, and the means to make further copies. This Universal Turing Machine would be capable of implementing any finite algorithm, given enough time and resources, because it need only be reprogrammed, not rebuilt, to perform the necessary processing.

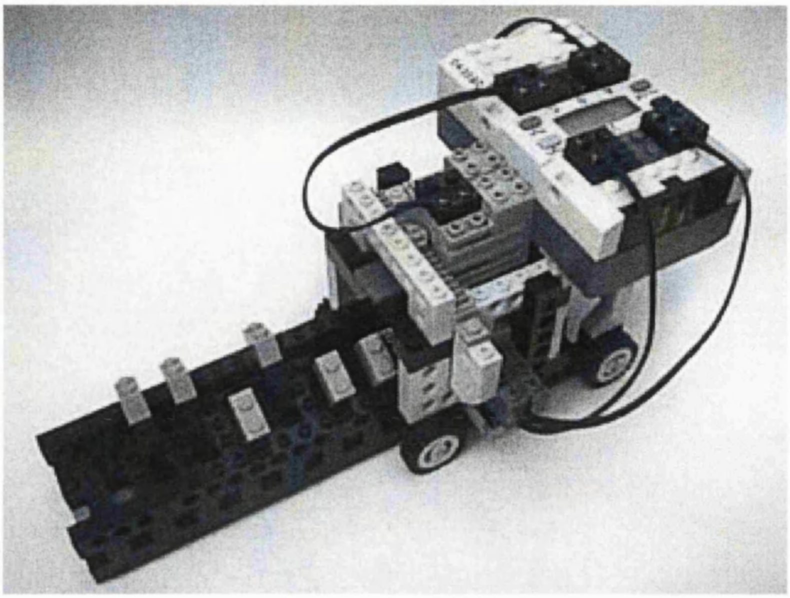


Figure 5. Lego Mindstroms™ Turing Machine

(Source: http://member.nifty.ne.jp/mindstorms/gallery/k0251.jpg).

It is useful to discuss the specification of the Turing Machine at this stage, as a prelude to subsequent discussion in later sections about adapting 'general' automata for urban simulations. The (hypothetical) Turing Machine consists of an infinitely long tape ruled off into sections. Each section of the tape contains one bit of information: a symbol with either a zero or a one. The Machine houses a scanning head (Figure 5) that is capable of being in any of an n set of configurations or states (in this example the set is binary). Time is discrete in the Universal Turing Machine's "world", it moves along in chunks as large or as small as one likes. Between time steps, the tape head examines its external world. Then, after consulting a rule table (see below), it considers the information that it encounters on the tape, as well as the current state of the tape head, and from this it determines its action in the next time step. The control mechanism on the tape head serves as a finite state machine (Sipper

1997). The tape head is the machine. It moves along the infinitely long tape; when the head reaches a section of tape, it can perform one of several actions (Casti 1997):

(a) Change the current state of the head to another state

(b) Retain the current state of the head

(c) Print `1` on the section of tape

(d) Print `0` on the section of tape

(e) Move left along the tape by one square

(f) Move right along the tape by one square

(g) Halt movement

From these simple rules and basic parameterizations, the Turing Machine is, hypothetically, capable of performing universal computation and thereby capable of self-reproduction.

Turing Machines were also associated with early consideration of Artificial Life. The Physical Church-Turing Hypothesis conjectured that a Universal Turing Machine could duplicate the functions of both mathematical machines and those of nature (Levy 1992).

Cellular automata

Cellular automata (CA) are a particular class of automata. CA share all of the characteristics of what we can term 'general automata', i.e., the automata discussed in the previous sections. However, CA have some important, unique, distinctions. First, CA are *parallel* processors rather than *serial* processors. In parallel processing, more than one particular process is active at any given time. In serial processing, on the other hand, one stage in the process is computed before the next starts; only one stage is active at any given time.

Second, the units that house individual automata are cellular in nature. Networks of connected CA can thus be understood to form a cellular lattice. The introduction of cells and lattices greatly extends the general automata framework. CA neighborhoods are defined in terms of localized areas of this lattice. The temporal evolution of cell states destroys the independence of initial cell states, instead prompting correlations between cells states at separated sites (Wolfram 1994). Consequently, CA can support a limited form of action-at-a-distance, over the course of their evolution.

Lattices are specified in one dimension in the simplest of CA (Figure 6), although lattices of any dimension can be specified (Figure 7, Figure 8, Figure 9). Lattices may be designed in a regular fashion, as grid squares or other combinations of regular shapes (hexagons and triangles are common). Irregular geometries may also be used, e.g., voronoi polygons and graphs. In common usage, lattices extend to infinite proportions within a given dimension. One-dimensional CA are designed to "loop"; two-dimensional CA are set to "wrap" as a torus (Figure 9).

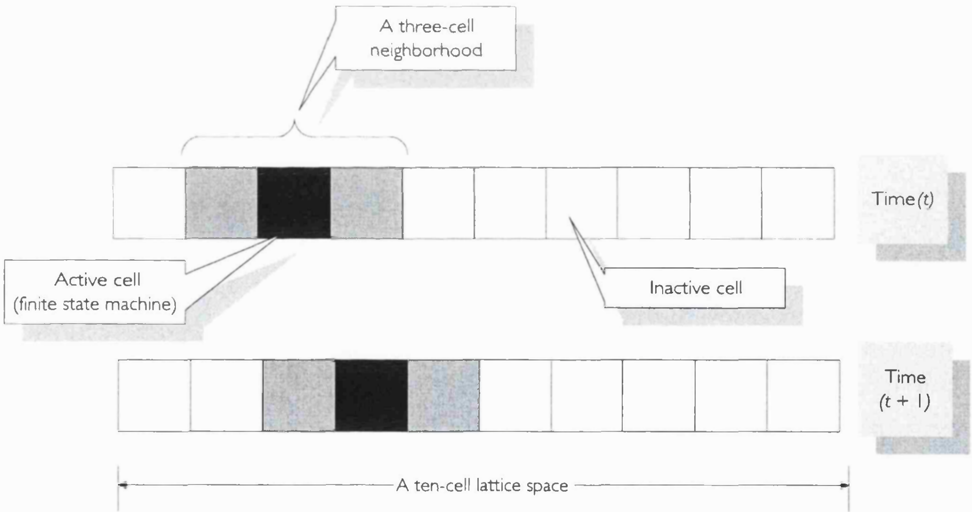


Figure 6. One-dimensional cellular automata

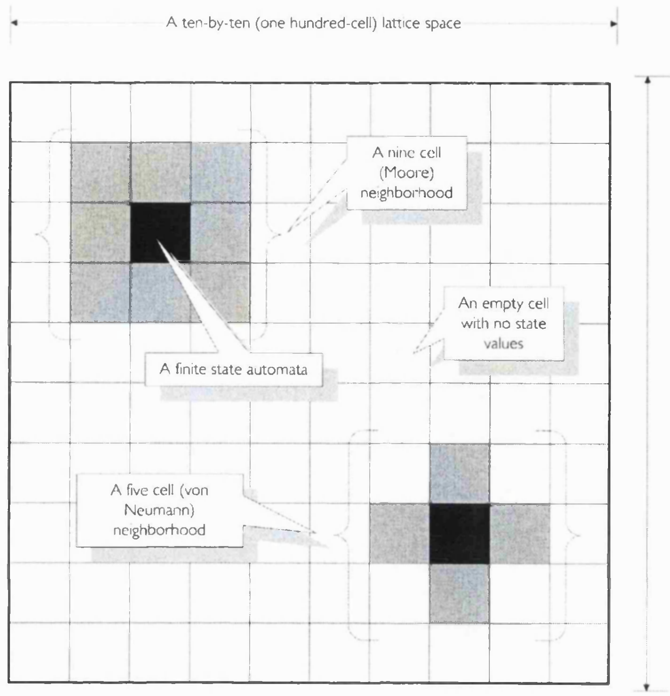


Figure 7. Two-dimensional cellular automata

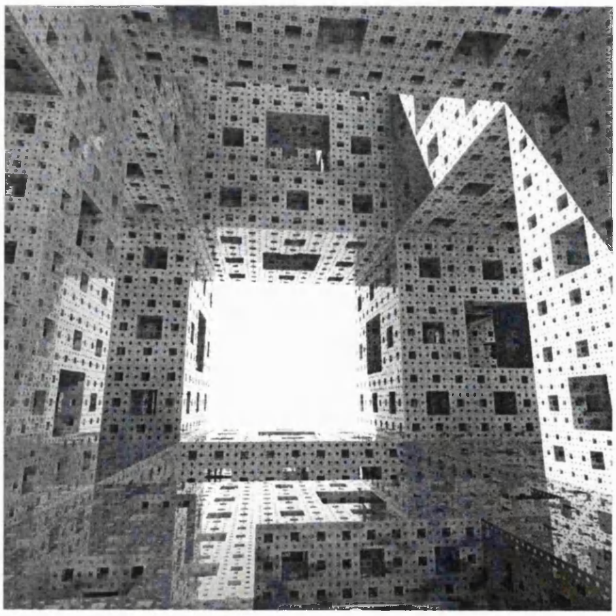


Figure 8. A three-dimensional Sierpinksi carpet

(Source: Paul Bourke, Centre for Astrophysics and Supercomputing, Swinburne University. http://astronomy.swin.edu.au/pbourke/fractals/gasket/).

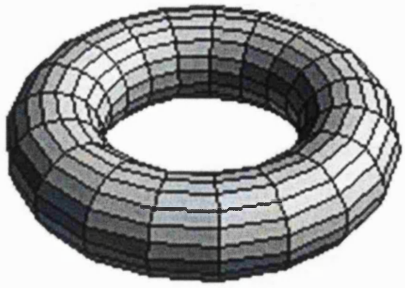


Figure 9. A torus lattice

CA have their origins in computing media, as is the case in the context of basic automata. CA were first devised by John von Neumann (the originator of game theory, and a pioneer in set theory, quantum mechanics, and the specification of electronic computers) and Stanislaw Ulam (who worked on Monte Carlo simulation and the hydrogen bomb (as part of the Manhattan Project with Edward Teller) and was influential in set theory and number theory) in the 1940s as a framework for investigating the logical underpinnings of life. "One can say that the "cellular" comes from Ulam and the "automata" comes from von Neumann" (Rucker 1999, p.69). Von Neumann and Ulam were interested in exploring whether the self-reproducing features of biological automata could be reduced to purely mathematical formulations—whether the forces governing reproduction could be reduced to logical rules (Sipper 1997). Here then, the connection between CA and A-Life is apparent.

One-dimensional cellular automata

One-dimensional CA are among the simpler CA. Figure 6 illustrates one-dimensional CA (the lattice extends in only one dimension) at time t and time $(t + 1)$ in their evolution. In this example, the lattice consists of ten independent automata (cells), organized as a one-dimensional space. Each discrete cell in this space can be in one of two states, either "empty" (depicted by the color white) or "full" (depicted by the color black). Each cell is driven by a transition rule table, which governs the state of the cells in each time-step. In the example presented in Figure 6, there are two time-steps in the life cycle of the CA. In the first time step, each cell draws input from its neighborhood (colored in grey), upon which an individual cell's behavior (transition calculation) can be based in the next time step. The transition rule picks out those cells in the lattice that are "full". The rule dictates that "full" cells take input from their two-cell neighborhood, both to the left and to the right. If that input contains information that neighboring cells in its right-hand side neighborhood are "empty", then the target cell transitions to "empty", while the cell to its right transitions to "full".

We could also formulate the CA using mathematical notation. The notation to describe a single cell is as follows.

$$S_{i,L,t+1} = f\left(S_{i,L,t}, I^{h}_{j,t}\right) \qquad \text{Eq xxxvi}$$

Where $S_{i,\,L,\,t+1}$ is the state of a given cell i, associated with a lattice L, at time (t + 1). f ( ) describes a functional relationship, where $S_{i,t}$ and the input from a neighborhood I (of neighborhood size h) in the vicinity of cell j at time t influence cell state transition in the next time step. The *entire* CA may be described in notation as:

$$\{S_{t+1}\} = f\left(\{S_t\} \cdot \{I^{h}_{t}\}\right) \qquad \text{Eq xxxvii}$$

In the above equation the notation is identical as before, except the letter S appears (without a subscript i), denoting the set of *all* states of cells in the CA, and $\{I^{h}_{t}\}$ refers to the set of *all* input neighborhoods.

The number of possible configurations of the lattice is X^n: the number of possible cell states (X), raised to the power of the number of cells in the lattice (n). For the example demonstrated in Figure 6, there are two possible cell states ("full" or "empty"), and ten cells in the lattice. That corresponds to 2^{10}, or 1024 possible configurations.

Two-dimensional cellular automata

Two-dimensional CA do not differ radically from one-dimensional CA in formulation; their lattice simply extends in an additional dimension. However, in terms of their simulation capacity, two-dimensional CA differ greatly from their one-dimensional counterparts. The example in Figure 7 would produce 2^{100}, or 1,267,650,600,228,229,401,496,703,205,376 possible configurations. Of course, intuitively, running two-dimensional CA models makes more sense than one dimension in many examples, including urban applications. The two most commonly defined neighborhood templates for a two-dimensional CA are the Moore neighborhood and the von Neumann neighborhood (Figure 7), although researchers have tinkered with neighborhood template sizes and configurations (Shi & Pang 2000).

Cellular automata as computing media

CA were developed, primarily, as computing media. Some famous experiments have been performed, examining the computational power of CA under simple specifications, and also developing primitive A-Life.

Wolfram's cellular automata classes

The mathematician Stephen Wolfram experimented heavily with CA, exploring the range of possible configurations that CA could evolve to with simple parameterizations and basic rule sets. Essentially, Wolfram was searching for the most minimalist CA that would be capable of universal computation. Wolfram's initial experiments involved one-dimensional CA. Wolfram discerned a broad typology of CA, consisting of four classes based on the dynamic behavior of CA and the patterns that they generated (see Wolfram 2002 for more details).

Class I CA: Evolve (or emerge) to limit points: fixed homogenous states that exhibit the maximum possible order both at local and global scales. In a CA with just two possible states—zero and one—this results in a CA that evolves to a condition where all cells have a value of zero.

Class II CA: Evolve to limit cycles: simple separated periodic structures that exhibit global order, but not of a maximal variety. Commonly the pattern of their evolution looks like a set of vertical stripes or cyclical 'railroad' patterns, when the patterns in the one-dimensional lattice are examined through time.

Class III CA: Evolve to chaotic aperiodic structures: patterns that exhibit maximal disorder at both local and global scales. The pattern that these CA generate need not be random, necessarily; often they are self-organizing.

Class IV CA: Evolve to complex structures, some of which are very long-lived. In some cases their complexity suggests that they may be capable of universal computation (Wolfram 1994).

Langton reversed the order of the classes in his description (Langton 1992), arguing that Class IV appear, in an evolutionary sense, before Class III: complexity leads to chaos.

The Game of Life

The mathematician John Horton Conway's Game of Life is perhaps the most famous instance of a two-dimensional CA (indeed, of any CA). Like Wolfram's one-dimensional CA, the Life CA model was developed by Conway to explore the simplest possible configuration for a universal computer. Although the Life CA were originally specified on a physical gaming board (actually, on the floor of Conway's office, later spreading into the corridor and common room of his college in Cambridge University), its computer recreations are among the first instances of A-Life. Conway spent some time tinkering with different parameterizations and rule sets, finally settling on the CA known as Life. Its specifications are very simple. Only two possible states are permitted in the Game: "alive" and "dead". The lattice of the CA is a square grid of infinite dimensions and the lattice turns in on itself to form a torus.

The neighborhoods of the Life CA are specified as Moore neighborhoods, consisting of nine cells. The transition rules are straightforward. There are three rules that govern dynamics ("life") in the game: birth, death, and survival. The birth rule specifies that a cell will be born (i.e., that it will transition from a state of "dead" to "alive") if it has three "alive" cells in its nine-cell neighborhood. Cells die (they transition from a state of “alive” to one of “dead”) from overcrowding between time steps if they have more than three live neighbors. Cells die by exposure if there are fewer than two live neighbors. The survivor rule specifies that a live cell should remain alive in the next time step if it has either two or three live cells in its neighborhood.

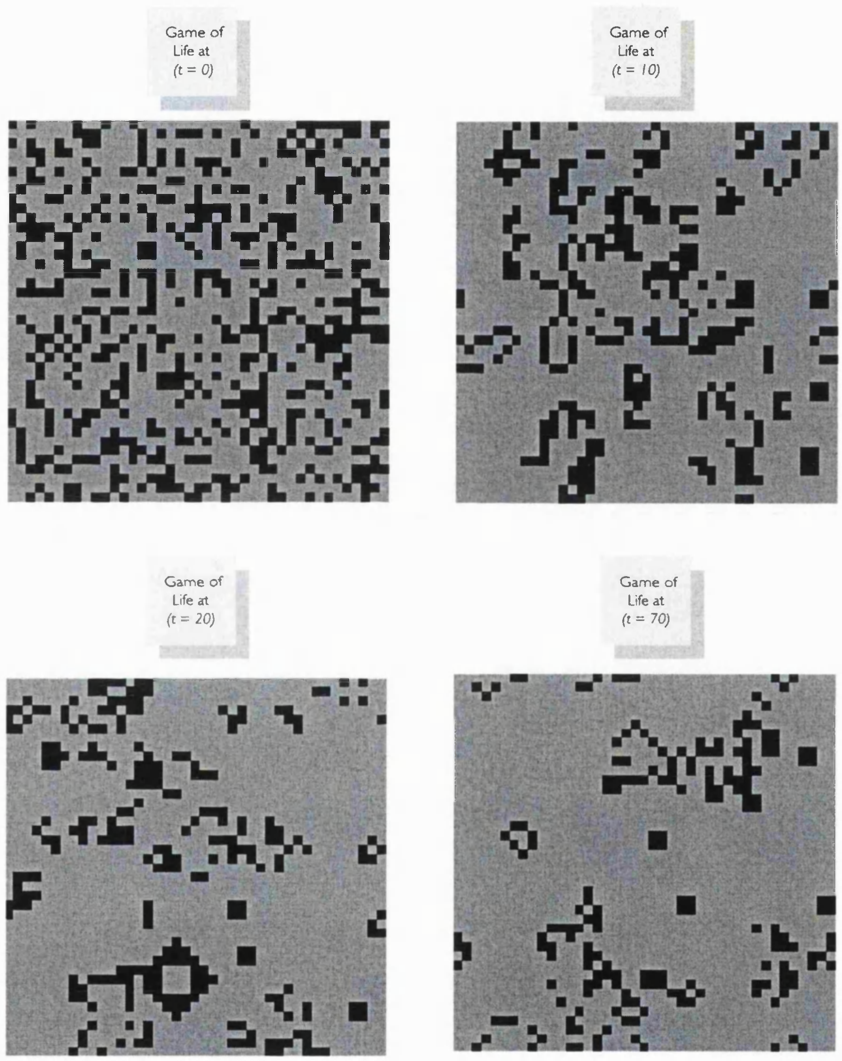


Figure 10. The Game of Life

(Black cells are “alive”, gray cells are “dead”)

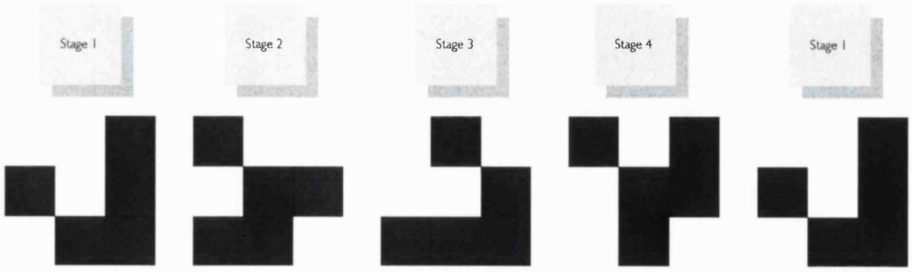


Figure 11. Glider evolution under the Life CA rules

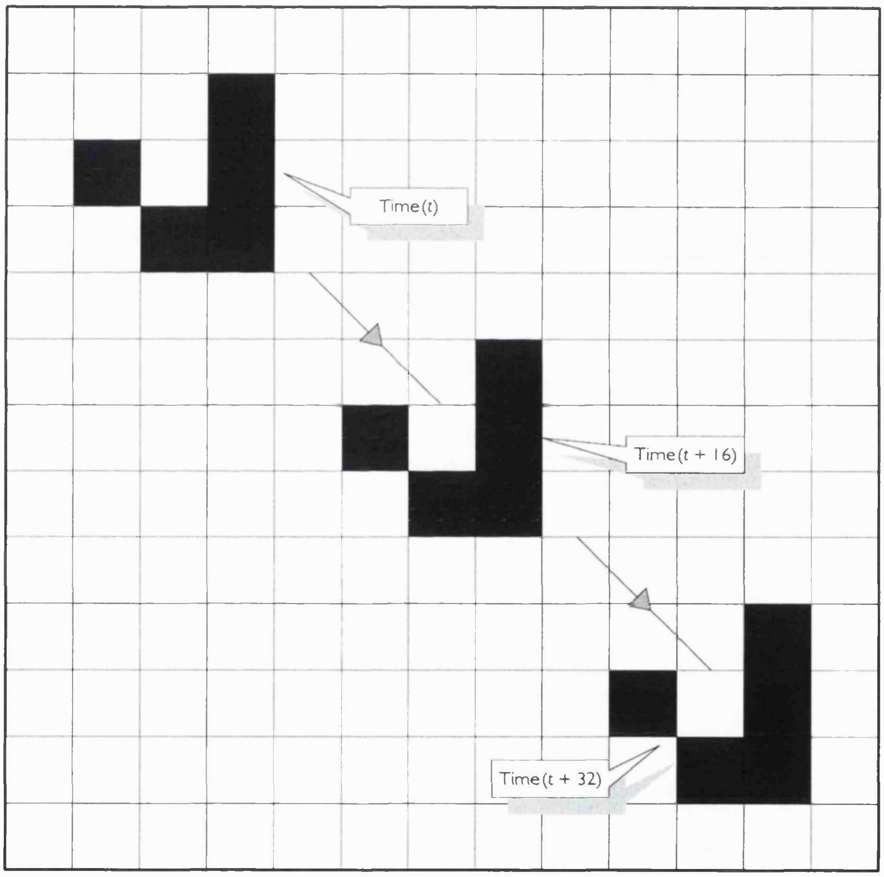


Figure 12. Glider movement

Once the game is run repeatedly with these simple parameterizations and rules governing play, persistent features begin to manifest themselves in the patterns that the game produces (Figure 10). In order to demonstrate that his CA was capable of universal computation, Conway needed an arrangement of cells within the lattice that could generate moving configurations of stable patterns. The search for such a configuration was opened to the public when Conway issued a challenge in Martin Gardner's column in the journal *Scientific American*; he described the problem and offered a cash prize for the first person to demonstrate the existence of such a configuration. The prize was claimed by R. Wilson Gosper at the Massachusetts Institute of Technology, whose team had coded a version of the Life CA into a computer and found 'glider guns' (Figure 11, Figure 12) that were capable of firing a steady stream of wandering gliders (Levy 1992). Essentially, the MIT team had demonstrated that the Life CA was capable of generating a machine that could, in turn, reproduce copies of itself that were as complicated in their structure.

Cellular automata as simulation tools

CA originated as computing media. They were originally designed as hypothetical constructs, specified with mathematical notation. With the advent of digital computers, themselves formulated on automata principles, automata came to be replicated in digital media as computer simulations. CA are excellent media for simulation, for a variety of reasons. We will discuss their specific benefits in the remainder of this section, but some general advantages can be identified. Fundamentally, CA provide most—if not all—of the essential building blocks for a model. Cells serve as storage units, bounding particular elements of the simulated system in discrete partitions. Lattices serve as larger arrays or assemblies of these individual units, specifying how they should network and relate to each other by means of neighborhoods. States can be employed to represent an almost limitless range of variables for characterizing system conditions. Transition rules can be used to formulate processes, impulses, and events that drive system dynamics over time. In many senses, several of these benefits remedy weaknesses inherent in conventional simulation methodologies, as discussed in Chapter 2.

Detail

Before the advent of computers, we could only study systems with large numbers of interacting mechanisms by assuming that individual elements exhibited a typical or average behavior. The overall behavior of the system was assumed to be the sum of these average behaviors. But, of course, many systems are non-linear, and here it may not be assumed that the aggregate constitutes the sum of the constituent parts (Holland 1995). This idea has carried through to urban systems simulation, where complacency in the substitutability of aggregate level data for detail has long proliferated. This approach has been well criticized; in particular, the consequences for the way in which we consider cities—as macro-structures that account for urban change—are especially unwelcome (Batty 1998). The result has been a generation of models and range of methodologies that are divorced from detail (Torrens 2000b). The capability of CA to handle fine-scale dynamics with computational efficiency has made them candidates for a new generation of detailed simulations. This is a logical progression from micro-simulation efforts in operational land-use and transport models (Wegener 1996), but with an explicit and welcome attention to *spatial* detail

Decentralization

Another advantage of CA, and their parallel processing approach in particular, is that they are highly decentralized. According to Resnick, the decentralized approach of CA is symptomatic and reflective of a very broad trend of decentralization—an "era of decentralization"—sweeping through society in general (Resnick 1994a). Up until recently, the world, and particularly our understanding of the world, was quite centralized. Central control was assumed to drive dynamics in many systems. The formation of groups of birds into flocks, and the organization of their flocking behavior, was assumed to be guided by a centralized 'leader bird' (Resnick 1997). Resistance to evolutionary theory—which still persists (Belluck 1999)—was based on the idea that the forces driving creation were a centralized, one-off event (Resnick 1996).

It is, perhaps, not unsurprising that a bias towards centralization exists. Many phenomena are, indeed, centralized. Also, people generally participate in social systems where power and authority are centralized. The experience of self and the mind as a singular entity also emphasizes individuality. However, the world is becoming increasingly decentralized. Organizations have become decentralized, distributing rights, responsibilities, ownership, and power from the top down. Decentralization is also pervasive in science. In physics, the domination of the Newtonian idea of a centralized and linear link between cause and effect is slowly being replaced by a decentralized approach that treats systems in terms of dynamic nonlinear interactions and feedback loops. In psychology the idea of a centralized mind has been challenged, most notably by Sigmund Freud, who proffered the idea of simultaneous existence of multiple perspectives and narratives co-existing in separate areas of the mind (Resnick 1994b).

More importantly for urban studies, *cities* are decentralizing. This is explored in considerable detail in the context of sprawl in Chapter 6. The old monocentric core and its orbital band of related settlement is rapidly giving way to polycentric city systems dispersed over a much less centralized network of relations and a spatial structure that is shaped by centrifugal forces, rather than centripetal activity.

This is not news; cities have always ebbed and lowed in this fashion. Some interesting and new phenomena are occurring on present-day urban fringes, however, that suggest new forms of decentralization. The incorporation of newly-formed suburban "townships" in a United States context is one example. These are, essentially, urban clusters drifting on the edge of the main urban mass, and many of these areas have begun to consolidate planning, management, and regulation of their urban environment in an independent fashion. In some cases, exclusionary practices have been noted, whereby the township "locks out" other potential incoming residents. Some authors have speculated that this actually contributes in a significant manner tp sprawl formation (Pendall, 1999).

Dynamics

As mentioned in Chapter 2, previous generations of spatial models have often misrepresented dynamics: models moved in 'jumps' from one time period to another. Often, these jumps spanned several years—a period in which much could change in a city. The weak representation of dynamics was a by-product of calibrating the models on cross-sectional data, often relying on information from the Census, which is produced on a ten-year cycle in most countries. Recent advances have seen models calibrated on longitudinal data, with time steps of as low as one year (Waddell 2000).

Nevertheless, the models still remain relatively static. They are updated rather than 'evolved' between time steps. CA models represent a significant advance in the treatment of time. The models are inherently dynamic, and importantly they are *interactive* in their dynamics. Time is still discrete in CA models—they still move in 'jumps'—but the jumps may be so small as to approximate real-time dynamics if necessary and if data permit. This is useful for modeling cities: CA are flexible enough to allow multiple timescales to be represented in the simulation. This is important in systems such as cities, where the lifecycle of interactive events—long-term economic cycles, daily commuting behavior, hour-by-hour social interaction, etc.—varies temporally. As we will see in Chapter 6, dynamics are very important in the context of understanding sprawl.

Function and form

A major advantage of CA is the equal attention that they afford to space, time, and system attributes. This united approach has the benefit of imposing a modeling framework that forces model-builders to consider the system that they are simulating in an *interactive* fashion, where a change in one element has profound effects on others (Batty 1997). This has obvious advantages for modeling geographic phenomena, where actions are intertwined continuously over space. Essentially, CA allow us to model function and form, pattern and process, simultaneously and in an interactive manner.

Spatiality

CA are particularly adept at dealing with spatial phenomena. As has already been explored, traditional modeling techniques have a tendency to abstract from spatial detail. CA, on the other hand, make implicit use of that spatial complexity (White et al. 1997). Additionally, CA are good at handling proximal space. If absolute (Cartesian) space provides an *a priori* frame of reference (site), and relative (Leibnitzian) space handles the relations between objects (situation), then proximal space is the space that connects the two, linking site and situation through the concept of the neighborhood (Couclelis 1997). "A neighborhood surrounds a localized item or place but it also embodies the notion of proximity to that place, which is a relation" (Couclelis 1997, p.170). Contemporary spatial models are generally implicit in their treatment of space (e.g., spatial interaction and gravity models, econometric models, location-allocation models, and core-periphery models) (Couclelis 1997). Yet, the operational procedures used to analyze cities are generally absolute (e.g., GIS and remote sensing). In this sense, CA hold many advantages over linear models of urban phenomena.

Affinity with geographic information systems and remote sensing

CA are commonly regarded as having a natural affinity with raster data (Couclelis 1997). They seem well suited to GIS and remotely sensed information. There are other commonalities between CA, GIS, and remote sensing (Torrens 2004b). Both CA and GIS organize space into discrete areal units (grids in CA, and polygons or grids in GIS). Also, CA and GIS represent attribute information in a layered fashion (themes in GIS and state-spaces in CA), and manipulate that information with operators (overlay techniques, for example, in GIS and transition rules in CA) (Wagner 1997). In many cases, state data can be arranged in a GIS or via remotely-sensed images before being introduced to the CA. Other authors have suggested a closer coupling of CA and GIS, suggesting that CA act as the 'analytical engine' for a GIS (Gimblett 2002; Wagner 1997). We will explore these issues in Chapter 5 when we consider Geographic Automata Systems.

Visualization

CA are, by their very nature, a highly visual environment for simulation. This has several advantages for urban modeling. The visual aspect helps to engage model users—users can interact visually with the model. The same advantage also applies to uses in education (Resnick 1997). CA can convey large amounts of information at once. Complex procedures, aggregated outcomes, statistical trends, and comparative measures can be presented visually and diagrammatically, enhancing the accessibility of the research. Because CA are visually dynamic, the evolution of the system can also be displayed as it changes over time, as will be illustrated with models of sprawl in Chapter 8.

Infusion of complexity theory

Treating cities as complex adaptive systems is an innovative approach to urban studies. As we have seen, many urban systems may be regarded as emergent in their behavior and complex in their organization. The complexity approach focuses on the 'grassroots' of the system—emphasizing the interaction among elements—without sacrificing a holistic perspective. The complexity approach also focuses model development on important issues such as the importance of historical (seed) conditions, feedback between subsystems, interaction, dynamics, noise and perturbations, etc. Of course, urban CA also offer the potential of illuminating our understanding of complex adaptive systems in general; the opportunity to inform debates in complexity studies is rich.

Agency, agents, and multi-agent systems

Agents are another class of automata, based on the general automata idea discussed earlier in this chapter. Agents resemble CA, but extend that framework with attributes borrowed largely from AI. Fundamentally, agents are automata. Agents are autonomous units that are capable of processing information and exchanging it with other agents or entities. The agent-based concept is very general, however, and there is no universal definition; there is not a set of components that are agreed upon, that can be used to demarcate what is agent-based and what is not. "The notion of an agent is meant to be a tool for analyzing systems, not an absolute characterization that divides the world into agents and non-agents" (Russell & Norvig 1995, pp. 33). Or, as Kohler (2000) puts it, agent-based modeling is more art than science.

As general as the term may be, we can lay down some foundation for defining agents. It is, perhaps, useful to divide the concept into agency, agents, and multi-agent systems (MAS).

Agency

Interpretations of agency are, for the most part, a function of the particular examples and applications that a designer had in mind when specifying an agent-based model

(Franklin & Graesser 1997). The literature on agent-based modeling identifies two broad categories of 'agency': weak and strong. These two classifications are a by-product of the division of research threads in agent-based research, between theoretical and experimental work (weak agency) and development work (strong agency).

The first approach—theoretical and experimental work—focuses on mechanisms that come from the interaction of autonomous agents in a simulation (Ferber 1999), such as self-organization and emergence. The second approach—development work—concentrates on creating distributed agent mechanisms that are capable of accomplishing tasks (Ferber 1999). The emphasis, in that approach, is on topics such as communications protocols, coordination of agent actions, etc. This 'strong' interpretation of agency includes 'weak' notions of agency, but is more commonly conceptualized or implemented using concepts normally applied to humans (Terna 1998), e.g., mentalistic notions such as knowledge, belief, intention, obligation, emotion, etc. Strong agency is closer to AI than weak agency.

Agents

Agents are best thought of as automata, first and foremost. However, they are a particularly sophisticated class of automaton, with a wide range of attributes that lend them more life-like (agent-like) qualities that their CA cousins do not possess.

Agents are *autonomous* entities. They are independent; their dynamics are governed without the influence of centralized control. Individual agents may interact with other agents or entities, but they remain autonomous.

Heterogeneity is another important characteristic of agents. This allows agent-based models to dispense with ideas of 'mean individuals', and permits the specification of autonomous, individual, agent entities. Groups may exist amid such structures, defined from the bottom-up as assemblies of independent units.

Agents are also *active*; they are *interactive* in the context of other entities. Whereas CA could be considered passive, agents are *operative* by comparison, in the sense that they exert independent influence in a simulation. CA cells, by contrast, serve as mere conduits for the information that passes through them. The additional components associated with agents make them much more active than their CA counterparts, as will be explained in the following paragraphs. In particular, the *intelligence* that is often attributed to agents sets them apart from other automata.

Agents are often designed to be *proactive* (Terna 1998). Agents in these instances act to realize a goal or set of goals (Franklin & Graesser 1997). Economic agents can be designed to satisfy utility goals; political agents might be designed with particular agendas to satisfy; and geographic agents could be created to follow a set of spatial paths. The agents described in Chapter 8 are designed to sprawl; those in Chapter 9 are designed to exercise residential mobility.

Individual agents are commonly specified with *perception.* This can be interpreted as an extension of the neighborhood concept in automata. Agents 'sense', or are 'aware', of their surroundings. However, agent perception is not constrained to geometric notions such as lattice arrangements or network paths, although agent-based models can be designed in that fashion. A more cognitive approach to perception is generally considered in agent-based models; agents are often endowed with a cognitive model of their 'world' and the ability to identify entities within it.

Agents are often specified with *bounded rationality.* The rational actor is a feature in many conventional social science methodologies. Rational actors are perfectly informed individuals, assumed to have infinite computing capacity, which maximize a fixed (non-evolving) utility function (Epstein & Axtell 1996). Rational actor ideas have long been criticized because they hold little resemblance to real-world actors (Epstein 1999). Agents are ideal vehicles for circumnavigating rational actor assumptions, because of their heterogeneity. Individual agents are commonly specified without global information; their perceptual model of their world is not perfect, nor is it complete.

Communication is one of the distinguishing properties of that separate agents from basic automata. General automata are capable of exchanging information with other automata; a target cell in a CA lattice takes input from other automata around it, as defined by its neighborhood. Agents are also capable of these types of exchanges, but can also communicate with other agents or entities in a simulation—they may query other agents, searching for a particular type of information and choosing to ignore extraneous details. In this sense, agents could be regarded as having a form of social functioning.

Agents are commonly designed with *mobility*. This is particularly important in simulation contexts (and particularly so in spatial simulation). Mobile agents are not confined to fixed locations in the spaces in which they exist. (Recall that CA are fixed in their lattices.) Agents can move around. Coupled with their interactive capabilities and their intelligence, this greatly expands the range of possibilities for designing agent-based simulations.

Adaptation is another characteristic that is often included in the design of agents. Like all automata, agents change states as they evolve, transitioning through a set of conditions. In the case of agent-based models, the ability for agents to change based on past experience often features in their design, thereby granting agents the capacity for memory-driven adaptation.

In terms of *time*, agents are continuous, temporally. They exist in a continually running process (Franklin & Graesser 1997), although time may be represented, in a simulation, in discrete packets.

Multi-agent systems

A multi-agent system (MAS) generally consists of a community of agents, situated in an environment. Community refers to the relationships between individual agents in the system, and these may be specified in a variety of ways, from simply reactive to cooperative. Generally, a set of relations are specified within a community, linking agents to other agents and objects in their environment (Ferber 1999). Environments are the spaces that house agents and support their activities.

MAS are used, in modeling contexts, as an experimental medium for running agent-based simulations. "The researcher employs a multi-agent system as if it were a miniature laboratory, moving individuals around, changing their behavior and modifying the environmental conditions" (Ferber 1999, pp. 37).

The rules of a MAS model run simulations, invoking the various functions of its constituent components and interaction among the various objects within the system.

Often, there may be evolving co-adaptive interactions among agents—and between agents and their environments—in a MAS (Kohler 2000).

Multi-agent systems as computing media

Like CA, MAS were developed in computer science contexts. However, unlike CA, MAS have their intellectual roots in more contemporary computing. The broad and contemporary basis for MAS tools accounts, in large part, for the flexibility of the approach. Just as CA are employed as computing media, so too are agents. MAS are used to develop AI, as vehicles for network computing, as the basis for animat development, and for experimenting with artificial swarm intelligence.

Agents and artificial intelligence

Most of the characteristics of MAS stem from classic AI, particularly in the context of translating the real world into agents' cognitive models and when representing agents' reasoning abilities. This is what AI is good at: symbolic reasoning, search mechanisms, fuzzy logic, inference routines, etc. (Kurzweil 1990; Kurzweil 1999). The design of heuristics is an important factor in developing agent-based AI. Rather than being simple reactive units, heuristics allow for the ascription of some internal initiative in agents, e.g., 'mental' or 'cognitive' reasoning.

Agents are excellent vehicles for AI design; similarly, MAS are used as artificial laboratories for testing AI, particularly in 'social' constructs. MAS are especially prevalent in Distributed Artificial Intelligence (DAI) work (Ferber 1999). DAI are systems composed of heterogeneous distributed parts (agents). They are generally used in solving complex problems where solutions based on local, bottom-up, approaches render a problem more tractable. Each agent in a DAI functions as an expert system. Expert systems consist of databases of information about a particular subject area, a set of rules for making inferences based on those databases, and an 'inference engine' for applying those rules to problem-solving in a systematic way (Kurzweil 1990).

Agents, bots, spiders, and Webcrawlers

MAS are also widely used in the context of network computing, as bots, spiders, and Webcrawlers. These three designations are essentially aliases for the same thing: agent-based programs running in a network—often the Internet—environment. They are an extension of client-server computing and are commonly specified as individual programs, dispatched via a network from a client computer to a remote server for execution. Spiders and Webcrawlers are commonly used in Internet contexts. They scour the Internet, updating search engine databases by indexing Web sites that they encounter (Pallman 1999). Spiders and Webcrawlers may also be termed as 'bots' (Leonard 1997). Bots are software robots, more commonly associated with network agents that maintain contact, directly, with a user.

Bots, spiders, and Webcrawlers are agents. They can also be considered as part of MAS, including the network environment in which they exist and other objects that they interact with, such as routers, servers, Web pages, files, databases, etc. They are also specified with various agent-based behaviors: protocol negotiation skills, access rights, scheduling abilities, data interpretation routines, etc.

Agent-based research is influenced by developments in network agent computing, and influences it in turn. Many MAS simulation environments are in fact implemented over networks, allowing model users to run distributed MAS experiments. NetLogo's Hubnet (Center for Connected Learning and Computer-Based Modeling 2003a)—part of the NetLogo (Center for Connected Learning and Computer-Based Modeling 2003b) simulation environment—is an example. Also, many MAS models, such as TRANSIMS (Barrett et al. 1999) and PARAMICS (Wylie et al. 1993), make use of distributed processing when running large multi-threaded simulations (Nagel & Rickert 2001; Torrens 2004a). Parallel processing over networks of machines such as Beowulf clusters is itself a MAS of network agents.

Agents and animats

Agents are used as computing media in the design and delivery of human-based AI. However, they have also been used in A-Life research, particularly in developing synthetic non-human life-forms in simulation environments. Indeed, A-Life research is one of the main reservoirs for agent-based modeling work.

MAS have been enthusiastically researched in the context of animat work. The term, 'animat', is a fusion of 'animal' and 'artifact' (Meyer & Guillot 1994). Animats are agents that descend from animals, metaphorically, but have a virtual existence (Ferber 1999). Animat research focuses on the design of life-like synthetic animals (often as A-Life): the specification of their locomotion, movement, and behavior; the design of interacting groups of animals; and the analysis of their interactive behavior in simulated experiments.

Boids are excellent examples of agent-based animats. Boids were developed by Crag Reynolds in the 1980s (Reynolds 1987), as a tool for mimicking the flocking behavior of birds (although they are more consistent with fish). The term, 'boid', is a cross between 'bird' and 'android'. Boids interact in a similar fashion to cells in Conway's Game of Life, in so much as they are specified with a minimal set of characteristics and simple rules, but are capable of supporting very complex phenomena.

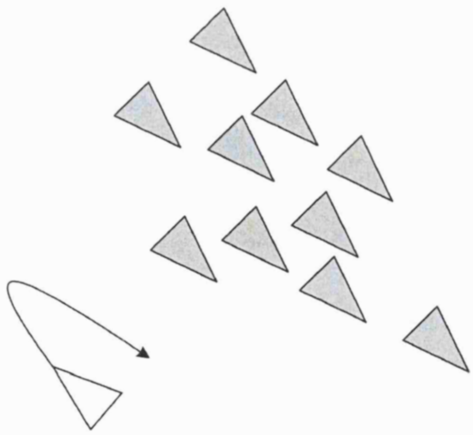


Figure 13. The Boids mass rule

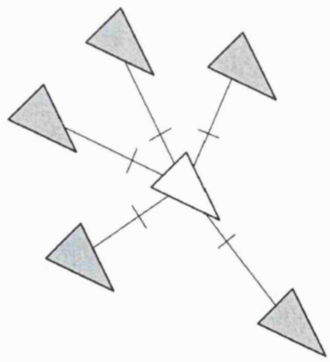


Figure 14. The Boids distance rule

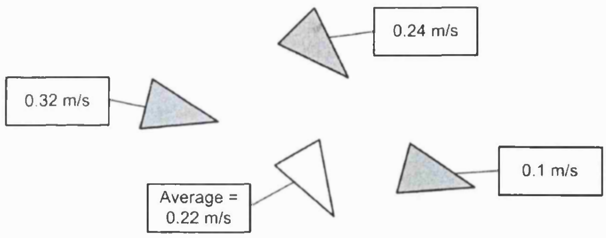


Figure 15. The Boids velocity rule

Individual boids are specified very simply: they have a direction, or heading, and a velocity. The collective behavior of boids is represented in a similarly lucid way. Three rules determine boid interactions. First, boids try to fly toward the center of mass of neighboring boids; the center of mass is commonly expressed as the average position of all the boids in a flock (Figure 13). Second, boids are designed to keep a small distance away from other boids (or obstacles, if they are included in the model). Essentially, this is a collision detection mechanism; boids are programmed to displace a set distance away from neighboring boids, within a distance threshold (Figure 14). Third, boids are configured to match the velocities of nearby boids, by speeding-up or slowing-down. This is generally achieved by matching an individual boid's velocity to the average velocity of nearby boids and then adding or subtracting speed, as may be required (Figure 15). As is the case with the CA-based Game of Life, a variety of emergent phenomena can be interpreted when boid simulations are run. Flocking behavior often emerges, with boids forming collective streams and swarms that move as a cohesive unit, despite the absence of a centralized control mechanism (Figure 16).

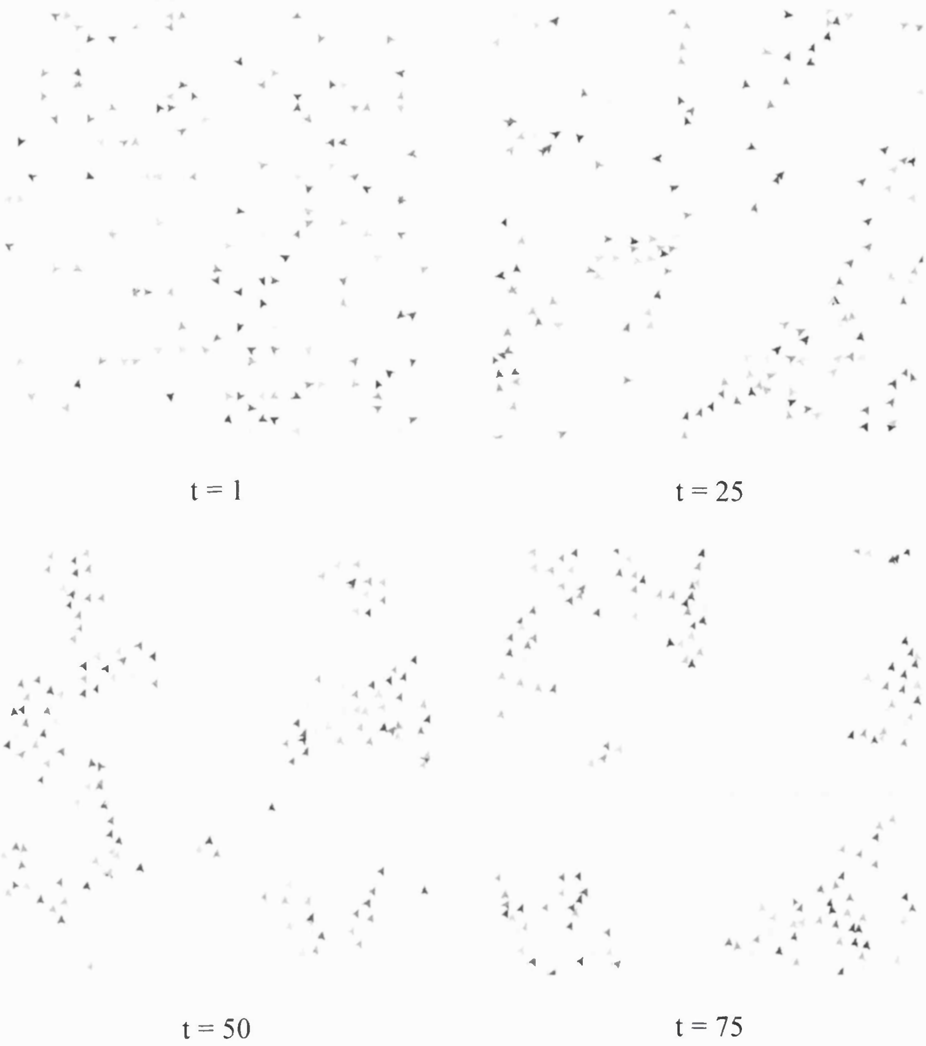


Figure 16. Boids flocking

Agents and swarm intelligence

The use of MAS designed along entomological principles is another popular domain of computing with agent-based media. For the most part, this work is focused on developing AI with MAS, based on ideas of 'swarm intelligence' from entomological work (Bonabeau et al. 1999). The idea is to fuse observed properties from social insect colonies—usually ants—with design principles from AI research. Biological and entomological metaphors, rules, and variables are used in the specification of MAS models, and simulations are run as decentralized problem-solving exercises.

Much of this work is based on analogies between agents and insects, and MAS and insect colonies. Many of the individual behaviors and phenomena associated with ants, for example, have parallels in other fields: finding food, building nests, responding to external challenges, etc. (Bonabeau *et al.* 1999). Ant-like models of urban systems have also been developed (Batty 2001b).

Multi-agent systems as simulation tools

MAS offer all of the potential associated with general automata and CA. However, they are significantly more advanced as simulation tools, when compared with other classes of automata. Despite their relative superiority as a modeling tool, MAS have not been as widely used in geographic research as CA, although there is no intuitive reason why this may be the case.

The main advantage of the MAS approach is its flexibility. This is particularly relevant in the context of using MAS as geographic models. MAS tools also have advantages for spatial simulation. As was mentioned earlier in this chapter, one of the characteristics of MAS that distinguishes them from CA is that agent automata are not confined to a lattice. This has significant advantages for simulation, as MAS can be specified within any environment. This can be a tessellated space, but that need not be the case.

This is related to another advantage of the MAS approach—the ability to specify agents with true fluid motion. CA cells cannot move within their lattices. The only form of movement supported in the CA framework is exchange: *information* is passed between neighboring cells, but those cells do not shift position. As was illustrated in the boids example, agents in MAS are capable of displacing, spatially, within their environments; they can move in different directions at varying velocities. This makes MAS very flexible, allowing for a wider variety of potential variables and parameters to be specified than would be possible in a CA framework.

This also has advantages for the specification of neighborhoods and agency in simulations. Neighborhoods can be specified in MAS using a variety of mechanisms.

It is relatively easy to introduce agent interactions mediated by space, networks, or combinations of structures. This is something that is traditionally difficult in mathematics, for example (Axtell 2000). It is also quite significant that MAS can support action-at-a-distance *directly*.

Another important advantage of the MAS approach to simulation is the broadly-developed rationale for equating agent units with real world entities. This is a legacy of the development of MAS in AI contexts and renders MAS inherently suited to simulating people and objects in very realistic ways.

Support for heterogeneity is also significant. Agent automata can be designed to represent any type of unit. Additionally, mechanisms can be introduced to form intuitive collections of individual units, from the bottom-up; agent individuals could be associated with a particular community, for example (as we will do in Chapter 9, for example). Heterogeneity also allows for the specification of agents with limited rationality (Axtell 2000), and this offers advantages over traditional approaches, which often assume fully rational individuals, if individuals are considered at all.

Finally, MAS offer additional advantages as open-ended simulation tools for problem-solving. As with CA, MAS simulations can be run over and over, with almost limitless opportunities for variation or alternative evolution; "executing the model—spinning it forward in time—is all that is necessary in order to "solve it"" (Axtell 2000, pp.2). An entire dynamic history of the system being studied can be mapped out.

Automata as urban simulation tools

Thus far, this chapter has described the development of automata—general automata, CA, and MAS—as computing media, and has discussed their use as general simulation tools. In the sections that follow, the use of automata as *urban* simulation tools will be discussed. The emphasis is placed on the adaptation of automata computing media for urban applications, with particular attention to the specification of various automata components and their generalization as urban entities and systems. First, CA models of urban systems are discussed, followed by MAS models. These sections serve as an extension of the computing media discussion, and as a prelude to subsequent chapters about Geographic Automata Systems and automata-based modeling of suburban sprawl. They also highlight the differences between CA and MAS as urban simulation tools, focusing on the use of MAS to model traffic, as opposed to the land-use basis of urban CA. The distinction is important—both methodologies offer quite different functionality for simulating cities, and as will be discussed in Chapter 5 and demonstrated in part three of the thesis, there is great potential for coupling the two. Indeed, merging both techniques offers considerable benefits for modeling sprawl in terms of the characteristics discussed in Chapter 6.

Urban systems as cellular automata

CA have been used quite widely in urban studies in recent years. This section provides a brief overview of urban CA applications. A more thorough review of models with particular relevance to sprawl will be presented in part two of the thesis.

The most common applications of CA models are to land development, urban growth, and land-use transition. In terms of land development, CA have been used to investigate the role of density constraints in development and the spatial distribution of growth activity (Batty et al. 1999b). Other models have characterized the development process as a profit calculation, mediated through space with the use of decision-making regimes from game theory (Wu & Webster 1998). CA have also been used to simulate development as a function of demand, supply, and potential (Batty 1998).

CA models have been applied to simulating urban growth processes, as well as specific forms of urban growth such as polycentricity (Wu 1998). CA have also been used to represent the evolution of urban form through growth cycles (Batty *et al.* 1999b). Growth has been modeled as proceeding from historically-identified 'seed' cells, using self-modifying transition rules that mimic the adaptability of cities over time (Clarke et al. 1997), and using predator-prey models (Batty et al. 1999a). White and Engelen (2000) have developed CA models that rely on exogenously-defined growth engines that reflect the position of cities in larger economic regions and economies.

CA have been quite widely applied to simulating land-use transition (Tobler 1970). Land-use dynamics have been modeled as a hierarchical process (White & Engelen

1997) and as a function of the rent-bidding power of individual sites and the externalities that they might produce (Webster et al. 1998). The role of inertia in land-use transition has also been investigated, as has the role of constraints such as accessibility (White 1998), density, and topological factors (Clarke *et al.* 1997).

Urban models are an abstraction, simplified versions of real world objects and phenomena that are used as laboratories for exploring ideas about how things work in cities. CA are no exception to this characterization. However, the basic CA, as defined by Ulam, von Neumann, Conway, and Wolfram (Poundstone 1985; Sipper 1997; Wolfram 1994) is not well-suited to urban applications; the framework is too simplified and constrained to represent real cities. To be successfully applied to the simulation of *urban* systems, it is necessary that CA be heavily modified from the formal parameterizations outlined earlier in this chapter. Indeed, quite radical modification is necessary before CA can approximate even a crude representation of an urban system. This often necessitates the introduction of additional components to add functionality to the basic CA design. The following sections discuss urban CA modifications in detail, referring to adaptation of cell-states, lattices, neighborhoods, time, and transition rules.

It is relatively easy to generalize the basic specification of CA to represent urban systems. The lattice on which a cellular automaton operates can be thought to represent any spatial realm: a city, a single floor in an office, an ecological habitat, etc. The CA lattice can be used to represent spatial structure of any description, such as road and rail networks, cadastral boundaries, delineations between ecosystems, etc. CA cells operate just like the pixels that comprise a television screen, except that each cell is capable of processing information, as well as visualizing its state. Cells can correspond to any zonal geography within a city: parcels of land, administrative boundaries, traffic analysis zones. In an urban context, the cell state can be made to represent any attribute of the urban environment, e.g., land-use (residential or commercial), density (high density or low density), land cover (forested or concrete), etc. Neighborhoods in urban CA represent spheres of influence or areas in which spatial interaction takes place, e.g., drainage basins, residential submarkets for housing, election districts. CA rules drive the dynamics of change in a model, and they can be devised to mirror how any phenomena in an urban system might operate, and can then be coded as algorithms within the simulation.

Cell-states

In basic CA, the cell-space is a closed environment. External events cannot influence dynamics within the CA. When CA are configured in this manner, there is no place for independent forces that might enter the model at a macro-scale (Couclelis 1985). Naturally, system closure makes little sense in the context of cities, where exogenous links and dependencies are numerous. To overcome this limitation, urban CA are often opened to outside influences, most commonly through constraints and algorithms applied to transition rules (see Clarke *et al.* 1997; Couclelis 1985; Semboloni 1997; White & Engelen 1997; White *et al.* 1997).

Cell states have also been reformulated in a hierarchical fashion in urban CA. The hierarchies are used to introduce the notion that state transition in urban contexts (land-use dynamics, for example) has a predilection toward pursuing fixed paths and proceeding in a sequential fashion (White & Engelen 1993).

In basic CA, cell-states have a certain level of homogeneity. Cells may adopt concurrent states (von Neumann's CA had 29 states, for example), but the data types describing those states have often been of the same form. Research into urban CA models has attempted to introduce a greater degree of flexibility into cell-state design by permitting cells to adopt concurrent states in a variety of data forms. For example, binary states—developable or not developable—appear alongside integer state descriptions (land-use categories, for example) and continuous values that correspond to various urban characteristics and properties such as land value and population counts (Wu 1998). Other authors have built CA models with 'fuzzy' states, representing proportional membership in a state class (Yeh & Li 2002). The introduction of cell state fixture (White & Engelen 1997) is an innovative concept. Here, a distinction is made between cell states that are 'fixed' and those that are 'functional'. In the context of land-use, we may regard sites that are generally exempt from the urban development process (such as water bodies) as 'fixed'; sites that are active in the development process (such as vacant lots) may be considered 'functional'. This idea is used later in this thesis, in the development of sprawl models.

Lattices

Basic two-dimensional CA are commonly defined on an infinite plane that is structured into a tessellated grid of regularly-spaced squares, or cells. Both the idea of an infinite spatial plane and that of a uniformly regular space are unrealistic for most urban contexts. CA used to study cities are often constrained to finite dimensions, with various tricks (such as buffers) for the treatment of edge effects (White *et al.* 1997). Recently, researchers have also experimented with three-dimensional lattices for urban CA, approaching a more realistic representation of the dimensionality of urban systems (Semboloni 2000). The idea of a *cellular* space for these models is also problematic for urban applications. Many features of cities are regular: some block configurations, building façades, internal floor plans, and many road networks. However, most objects in cities are not regular, and are certainly not square in shape. In a bid to afford urban CA models a greater degree of realism, researchers have introduced a variety of irregular lattice structures into the CA framework, re-specifying lattices as a graph (O'Sullivan 2001) or as Voronoi polygons (Shi & Pang 2000), for example.

Neighborhoods

The strict formalism of basic CA provides for a very limited specification of neighborhoods of influence. In basic CA, a neighborhood consists of an individual cell itself, as well as a set of adjacent cells at some distance from the target cell. In basic two-dimensional CA, there are two popular neighborhood configurations: the Moore neighborhood of the eight cells that form a square around a cell, and the von Neumann neighborhood of the four directly adjacent cells that comprise a cross centered on a cell (Figure 7). The rigidity of basic CA neighborhoods suppresses direct action-at-a-distance (although action does propagate, indirectly, across distances as the simulations evolve). In the urban world, neighborhoods of influence vary significantly and, more often than not, they fail to fit into the neat typology offered by basic CA. Social interaction, for example, can operate between adjacent properties, as well as functioning on a citywide scale. In Chapter 5 we introduce a new framework for representing neighborhoods and this allows for the specification of multiple neighborhoods concurrently in the models described in Chapter 9, in particular.

It is not surprising that neighborhood parameters have been modified in urban CA. Distance-decay effects have been introduced, often as weights applied to neighborhoods in transition calculations. Also, neighborhoods have been extended to comprise larger spaces (White & Engelen 1997; White *et al.* 1997). The introduction of 'fixed' and 'functional' cell-states can also serve as a proxy for asymmetric neighborhoods, as 'fixed' cells can serve to remove areas of the neighborhood from a transition calculation.

Time

The temporal space of CA is also an issue of concern for developers of urban CA models. In the basic CA framework, time is discrete and cells are made to evolve synchronously between time steps. Transition rules are applied uniformly: all cells are updated at the same time. The urban equivalent of this would be a situation in which all events are resolved simultaneously. Researchers using CA for urban applications have experimented with asynchronous cell-update (Portugali 2000), attempting to circumnavigate the universal treatment of time (although, the ramifications of using asynchronous update are not yet fully understood in urban research). Others have tinkered with the cell-state transition process so that it is only partly sequential, fashioned in dependence of exogenous constraints (White & Engelen 1993, 1997; White *et al.* 1997).

Transition rules

Transition functions serve as the algorithms that code real-world behavior into an artificial CA world. In the context of urban CA, transition rules are responsible for explaining how cities work. As we have already seen, transition rules have been opened up to exogenous links, permitting urban CA to function as quasi-open systems. Also, they have been reformulated as probabilistic expressions, a departure from the deterministic specifications of strict CA. In this manner, rules can be made conditional upon a probability, introducing an element of randomness, or 'noise', into the model. Also, probabilistic rules can be made dependent on other rules formulated within the model, reflecting the idea that urban systems operate as a tangled web of co-dependent sub-systems and phenomena. Weights for these rules have been calibrated against regression models (Yeh 1998) and artificial neural networks (Li & Yeh 2002), for example.

Self-modification, an idea not unlike evolution or mutation, has been used to expand CA functionality for urban purposes. Based on the idea of the genetic algorithm in computer science (Mitchell 1998), transition functions are allowed to change via feedback mechanisms as a CA model develops. In this sense, the rules 'evolve' (perhaps to some level of fitness or optimal efficiency) in reaction to the problem space of the model, adapting over time as the CA progress iteratively.

Transition rule formulation has also absorbed other simulation techniques from urban modeling, particularly ideas from regional science and econometrics (Torrens 2000b). Economic principles, such as utility maximization, have also been woven into transition rules (Webster & Wu 1998; Webster & Wu 1999a, b; Wu 1998; Wu & Webster 1998), as have accessibility algorithms based on spatial interaction (White *et al.* 1997). However, CA models have been slow to adopt explicitly geographic or urban theory as a basis for transition rule formulation, missing a valuable opportunity to infuse the models with a solid theoretical foundation (Torrens & O'Sullivan 2000a). This point is one of the premises for developing the GAS framework discussed in Chapter 5.

Urban systems as multi-agent systems

Compared with CA tools, MAS are not as popularly used in urban research, despite their added flexibility and arguable superiority as simulation tools. This is, perhaps, partly due to the large abundance of literature and tools in agent-based modeling, and the generality of the concept.

Most 'geographical' MAS simulation work has been researched in fields outside of geography, particularly in other social sciences (Epstein & Axtell 1996). However, such models are generally developed with very weak spatial representations and functionality. There is a large literature on 'geoagents'. This body of work is concerned with designing interface tools for geographical software—often, GIS (Camara & Raper 1999). Geoagents in these contexts manifest as buttons or widgets on Graphical User Interfaces (GUIs), designed to call up data or execute a particular command. They are not simulation tools. Much work has been done in physics on MAS models of movement, but for the most part, 'human' applications have been restricted to applications of motion in confined spaces, e.g., modeling escape panic in a room or building (Helbing et al. 2000; Vicsek 2001).

There is a good deal of misunderstanding of MAS in the urban literature, with CA and MAS being widely confused. MAS are often misinterpreted as 'mobile CA'. Essentially, such models are CA in design, but with some form of 'agency' attributed to the cells, usually in the form of some anthropogenic state variables; the cells are not mobile, however, and there are generally no transition rules governing motion, or neighborhood conventions to facilitate mobile entities (Benenson & Torrens 2004a).

Traffic dynamics—of vehicles and pedestrians—is one of the few areas of urban studies in which MAS models have been developed and applied. Here, also, there is some confusion between CA and MAS. In fact, most of these traffic models are specified as CA. However, they do make explicit reference to motion, with parameters designed to imitate movement.

Vehicular traffic

The agent-based approach allows for 'microscopic' traffic modeling, with individual vehicles being simulated as independent entities, and permits for the simulation of interactions between those vehicles along simulated roads. Often, recognizable traffic conditions, such as congestion (Nagel & Schrekenberg 1995), emerge from these interactions and in many cases the models provide mechanisms for simulated entities to react and adapt to these conditions as they evolve in near real-time.

For the most part, automata-based traffic models are developed in much the same way as one-dimensional CA. The spatial or network structure of the traffic environments that are being simulated are encoded into the model as lattices. Simulated entities are designed with various characteristics that enable them to function in a manner similar to their real world counterparts. Neighborhoods of influence designate the spatial domain of interactions between entities in the simulation. Some form of internal clock is introduced into the model, allowing for dynamic action. Conditional rules and calculations are also included in the model, describing how modeled entities should perceive their simulated environment, react to changes in their own state descriptions, react to other entities in the simulation, and react to changes in their environment. However, individual cells are assumed to be vehicles; their 'motion' is diffused along the lattice, thereby mimicking movement. Also, there is generally explicit consideration of human decision-making—driver behavior. In this sense, they are closer to MAS agents than to CA cells.

Spatial topology

In automata traffic models, roads are treated, to some extent, as the environment for MAS. Roads are encoded in familiar ways: lattice nodes represent road junctions and links represent roads that connect those junctions. Additional detailed topology may also be introduced. In the TRANSIMS model, for example, land-use and connectivity data, intersections (signs and signals), activity locations, parking, transit stops, and route paths are also encoded into the topology of the model (Barrett *et al.* 1999).

Queues are used to represent the vehicles that travel along a link. Queues are commonly encoded as one-dimensional lattices (or parallel lattices where multi-lane roads are represented), with each cell in the lattice represented as a cellular automaton. Where models are developed for experimental purposes, such as studying the formation of congestion in an abstract sense, queues may be coded with periodic boundary conditions, i.e., traffic moves in a continuous loop through the queue (Rickert et al. 1996). Various characteristics can be associated with the cells that form a queue, e.g., length, flow capacity, free flow velocity, free flow travel time, etc. (Cetin et al. 2001). In this way, automata cells and lattices can be used to build realistic traffic environments.

Entity descriptions

One of the great advantages of the MAS approach is that it allows simulated entities to be represented as atomic objects. Whereas spatial interaction models can represent aggregate flows of traffic, agent models represent the individual particles that comprise that flow: individual cars and trucks and their drivers for vehicular traffic, and individual walkers for pedestrian traffic. In most of the published microscopic traffic models, vehicles are encoded as individual agent units of 7.5 meters in length, which is the length of a car plus the gap between cars in a jam (Wagner et al. 1997).

Simulated entities can be afforded a rich range of state descriptors denoting their characteristics. There is no need for 'mean individual' descriptions; entities can be represented as true individuals. In the PARAMICS model, for example, individual vehicles are encoded with state variables that represent a vehicle's length, maximum acceleration and deceleration, cornering speed, desired destination, preferred traveling speed, current position, and current direction (Wylie *et al.* 1993). The TRANSIMS model is also capable of representing much of that information but adds even more detail to the description of vehicles, including a record of the household to which the vehicle belongs, the initial network location of the vehicle, and a vehicle classification from a 23 type scheme (Barrett *et al.* 1999).

Neighborhood definitions

Neighborhoods of influence can be defined for individual vehicle-automata in a simulation. These neighborhoods represent the range of influence for interaction between modeled entities. Neighborhoods are used to model drivers' "perception" of traffic conditions around them, such as the buffer between adjacent cars in the same lane, or gap opportunities for merging traffic at junctions (Wylie *et al.* 1993). Sophisticated neighborhood functions may also be introduced to facilitate lane-changing decisions; for example, how far to look ahead or behind in the same lane and how far to look ahead in adjacent lanes before switching position (Rickert *et al.* 1996). Neighborhoods can be defined in static terms, e.g., occupancy of five cells ahead or in front of a vehicle (Barrett *et al.* 1999). Or, alternatively, neighborhoods can be related to other dynamic characteristics of the model, such as the velocity of a moving vehicle (Rickert *et al.* 1996).

Time

The ability to encode dynamic relations in a simulation model is another advantage of the agent approach for traffic modeling, where users are often interested in moment-by-moment traffic dynamics for systems of interest. There are two ways in which the approach is particularly innovative in relation to its treatment of time: temporal resolution and parallel update. Traffic applications of MAS-style modeling are among some of the most fine-scaled simulations, in terms of temporal resolution, in urban studies. This is partly because they need to be—the reaction time of drivers is on the order of one second (Wagner *et al.* 1997)—and is partly a function of the incredible computing power available to compile and run these models. Further advantages stem from the synchronous nature of transition rules in the models. In keeping with automata-based principles, MAS traffic models are often updated in parallel (on supercomputers, clustered processors, or networks of machines); transition rules are applied to modeled entities and their states are altered in unison, throughout the simulation. When coupled with a fine-scale temporal resolution, this enables the simulated system to 'evolve' dynamically, analogous to real world conditions. In this sense, traffic models can now be run, in many cases, in near-real-time for medium size cities. In addition, the reaction of individual vehicles to *evolving* traffic conditions (accidents, congestion, detours) can be simulated dynamically.

Rules

Various transition rules can be used to characterize the behavior of vehicles, and their drivers, in automata-based traffic simulations. Of course, it would be a daunting task to formulate rules to mimic the full range of driving behaviors, so model developers focus on a minimal set of rules instead (Wagner *et al.* 1997). Just as in complexity studies, traffic simulations are designed with a few simple rules and it is hoped that more complex behaviors will emerge through the myriad application of those rules between many interacting entities.

MAS traffic models are noteworthy in their attention to rules of *movement*. Transition rules are formulated to simulate acceleration and braking as a function of various vehicle characteristics (speed, maximum velocity, target speed) and conditions in a vehicle's neighborhood (type of road, perceived traffic conditions ahead, gap to the next car) (Wagner *et al.* 1997; Wylie *et al.* 1993). In some instances, random acceleration and deceleration functions are also introduced to mimic erratic movement (Rickert *et al.* 1996). Rules for collision-avoidance have also been introduced. Other rules have been developed to simulate signal stopping behavior (Barrett *et al.* 1999) and traffic movement at junctions, with "gap acceptance" functions that determine how long drivers must wait at a junction before they can proceed (Wylie *et al.* 1993). In models where collections of vehicles are simulated as traffic queues (Barrett *et al.* 1999; Cetin *et al.* 2001; Rickert *et al.* 1996), entrance and departure from vehicle queues can also be simulated, with vehicles leaving a queue freeing up space on a link, allowing another driver to join the queue.

Quite elaborate rules have also been devised to simulate lane-changing behavior. In this sense, automata models resemble traditional queuing lane models. However, traditional queuing lane models are not truly multi-lane in their design (Cetin *et al.* 2001); they *approximate* multiple lanes by switching the positional order of vehicles to make it appear as if passing has occurred. In automata models, however, parallel lattices can be constructed adjacent to each other, facilitating the simulation of movement between lanes. Lane-changing rules in automata traffic models can the simulate exchanges of vehicles between lanes as a function of a variety of factors, including the number of empty sites in a vehicle's neighborhood ahead in the same lane, ahead in adjacent lanes, and behind in adjacent lanes; velocity; and hindrance in the current lane (Rickert *et al.* 1996; Wagner *et al.* 1997).

Pedestrian traffic

Traditionally, pedestrian traffic has been comparatively ignored by transport modelers. There are many reasons why this may have been the case (Batty 2001a). To a certain degree, pedestrian traffic has been overshadowed by vehicular traffic as an area of applied research. The demand for vehicular transport, at least in contemporary metropolitan areas in developed countries, out-paces that for pedestrian modes of travel by a significant margin. Likewise, the multitude of problems—environmental, health, social justice—associated with vehicular transport overshadows those ties to pedestrian travel. Scale issues also factor into the relative favor afforded vehicular transport. The range of movement permitted by vehicular transport, and the associated scale of its influence, are far greater than that of pedestrian movement. Vehicular transport problems are also, to some extent, more tractable than pedestrian transport problems (Batty 2001a), partly because of the aforementioned scaling issues, and partly because of the greater attention paid to vehicles and the wider array of modeling techniques that are available.

However, in recent years, the landscape for pedestrian transport research has improved considerably. This is partly a response to shifts in the political agenda in relation to transport, particularly heightened awareness of sustainability in urban environments and modes of transport. As in other area of urban modeling, the field has also benefited from innovations in the research landscape, such as those discussed in Chapter 2. Together, these advances have enabled the development of innovative, microscopic, agent-based pedestrian simulations in which the activity schedules and second-by-second movement and interactions of individual walkers are simulated, sometimes for large crowds of pedestrians in whole districts of a city. These models enable applied work to be performed that had either been previously intractable or not imagined at all.

Pedestrian traffic modeling is, in many respects, a far more complex simulation problem than vehicle traffic modeling. This is particularly true at 'microscopic' levels. The scale of observation can often be much finer for pedestrian modeling, simply because the 'footprint' of a pedestrian is generally much smaller than that of a vehicle. Furthermore, the behavior of pedestrians is not as constrained as that of vehicles on roads. There are generally many more paths available to pedestrians when compared to vehicles. Pedestrians are also much less limited in their range of movement; they can, for example, perform side-step and about-face maneuvers. Pedestrians are not generally constrained by rules of the road; they can, for example, ignore crossing rules at intersections by jaywalking. Finally, pedestrians themselves, as well as their behavior, are much more varied than vehicles, at least in a general sense. Despite the age, gender, social, cultural, and health characteristics of various drivers, most cars behave in a relatively similar manner on the road. That is not true of pedestrian walkers.

For these reasons, agent-based techniques are ideally suited to modeling pedestrian traffic. The comparative flexibility of MAS tools compared to CA, with respect to representing movement and interaction, makes them ideal for representing complex adaptive phenomena like pedestrian crowds. As with vehicles, most pedestrian traffic models are, fundamentally, CA by design. However, they closely approximate MAS by mimicking movement and agency.

Entities

Generally, agent-based pedestrian traffic models provide for the representation of two types of entities: agent-pedestrians and pedestrian obstacles in the built environment. The simulated vehicle drivers discussed in the previous section could have various demographic and socioeconomic state variables associated with them. This is also the case with pedestrian traffic models, where simulated agent-pedestrians are often attributed various life-like characteristics to help shape their movement behavior and to populate the models with agents that are likely to behave in a diverse fashion (Haklay et al. 2001). Other characteristics of relevance to traffic modeling can be observed as agents move within the simulation, e.g., position, direction, time in the system, movement, states, etc., and this has close analogies with other pedestrian flow modeling approaches (Hoogendorn & Bovy 2002).

State variables can also be ascribed to various entities used to represent the physical environment in which pedestrian agents interact, both in terms of attraction features (buildings, shops, etc.) and potential obstacles (street furniture, walls, road signs, etc.) (Kerridge et al. 2001).

Spatial topology

Invariably, grid-based lattices are used to represent the spatial topology of the environments in which agents interact, as is the case in vehicle traffic models. Of course, a finer resolution of representation is often necessary for pedestrian models; in some instances grid squares have been used to represent spaces of 750mm in size. Various features of the built environment—building outlines, land-uses, divisions between sidewalks and roads, network data, gateways and waypoints, etc.—may be embedded into this typology, either as raster or vector data. Also, various representations of street and building layouts can be altered in the model structure to allow for the evaluation of planning and design issues relating to pedestrian movement.

Time

As in most MAS-style models, time is generally discrete in pedestrian traffic simulations, proceeding in packets of change, commonly designed at very fine temporal scales. In the PEDFLOW model (Kerridge *et al.* 2001), for example, time is divided into slots of one tenth of a second in duration.

Neighborhoods

Various neighborhood functions may be introduced to determine pedestrian agents' "awareness" of conditions in the environments surrounding them, both for the detection of targets and potential obstacles and the determination of avenues for collision-avoidance. In their models of agent-based shoppers, Turner and Penn (2002) specify agents with neighborhoods derived from their lines-of-sight. In the STREETS model (Haklay *et al.* 2001), agents "look" in up to five directions in their immediate vicinity to determine where the most space is available for movement. In the PEDFLOW model (Kerridge *et al.* 2001), agents are equipped with three neighborhood filters: a "static awareness" function that determines how far ahead an agent can "see"; a "preferred gap size" that governs the smallest space a pedestrian is willing to move into; and a "personal space" function that sets the amount of buffering space a pedestrian would like to maintain around itself. These neighborhood functions provide simulated pedestrians with the spatial "cognition" necessary for realistic movement patterns.

Rules

It has been noted that, as is the case with vehicular traffic movement, there is an almost limitless array of behaviors and factors that contribute to the movement dynamics of a pedestrian crowd. Nevertheless, pedestrian movement is surprisingly predictable. Despite the apparent chaos of crowd dynamics, certain regularities can be observed and these can be used to formulate transition rules that determine movement behavior in agent-based simulations. Papers by Helbing and colleagues (Helbing *et al.* 2000; Helbing et al. 2001) detail several of these regularities. Pedestrians usually pursue the fastest route to a target destination and prefer to travel at the most comfortable walking speed. Pedestrians also like to maintain a buffering distance from other pedestrians and obstacles. Automatic behaviors may also be observed in certain situations, e.g., when entering congested doorways or side-stepping obstacles. Also, at a more macro-level, gas- or fluid-like qualities may be attributed to pedestrian crowds at certain densities, and similarities with granular flows have also been noted. These observations may be used to formulate transition rules governing the speed and movement of pedestrian agents in traffic simulations.

In the PEDFLOW model (Kerridge *et al.* 2001), for example, the speed of pedestrian agent movement is determined by factoring in the time period over which an agent occupies a grid square, proportional to its own walking speed or that of other pedestrian agents in its neighborhood. In the STREETS model (Haklay *et al.* 2001), pedestrian agents are endowed with maximum walking speed attributes and categorical variables that characterize their speed at any given moment (e.g., "stuck", "standing", "moving") and these variables are used to determine the speed at which a simulated pedestrian walks.

Agent movement—way-finding and navigation—is determined by rules that are analogous to those for vehicular traffic. Activity models determine the overall movement schedules of agents, and determine target destinations or waypoints. (Agents may decide to adhere to those schedules or wander from pre-assigned targets.) Various navigation rules then determine how agents navigate to those destinations within their simulated environments, reacting to and interacting with the emerging dynamics within the simulated system. In the STREETS model (Haklay *et al.* 2001), for example, various "helmsman" rules are used to mediate between an agent's "best heading" and its desired target, while "navigator" rules maintain agents' overall heading in relation to targets. On a more micro-scale, rules are often introduced to determine how pedestrian agents should react to evolving conditions in their immediate surroundings: to determine step-by-step movement and collision detection. In the STREETS model, a "walkability" calculation is performed to assess whether enough space exists ahead of an agent for it to proceed along its route. Agents then move to grid squares with the most "walkability". In the PEDFLOW model (Kerridge *et al.* 2001), agents perform similar calculations in relation to their neighborhoods, distinguishing between observed entities in that neighborhood (other pedestrian agents, goal points, stationary obstacles, buildings, and curbs). Agents calculate the distance to those objects and then execute one of four actions to proceed: go straight ahead, go diagonally left or right, move to the side (a choice parameter determines which side they favor), or remain where they are. Using these rules, pedestrian agents can be designed, choreographically, to mimic the movement patterns of real walkers, both at an individual scale and as a crowd.

Multi-agent systems, for all their advantages, actually pale in some areas when compared with traditional approaches. The emphasis on *individual* agents, for example, is at once an advantage and a weakness. It has advantages in terms of resolution and representation, offering a new perspective that computational models of this nature may have shied away from in the past, particularly in massively interactive contexts. Yet, the focus on the individual is also a weakness—representation of groups is sometimes problematic in a multi-agent context; ironically, this is what techniques such as micro-simulation handle well.

Conclusions

The main thread of this thesis began in Chapter 2 with a general overview of 'traditional' urban model methodology, and a critique of those techniques. Chapter 3 described recent developments in the geographical sciences and fields outside geography that have provided the basis for a paradigm shift, of sorts, in the field of urban simulation. Catalyzed by those developments researchers in geography and urban studies have begun to construct a 'new wave' of urban models, centered on automata-based tools. Two classes of automata have been particularly influential—cellular automata (CA) and multi-agent systems (MAS).

Both CA and MAS have origins in the early development of digital computing, and both are based around general automata principles. The general automata was introduced at the start of this chapter, followed by discussion of CA and MAS as computing media and simulation media. It is important to remember that CA and MAS *have* origins as computing media. The discussion of their use as *urban* simulation tools described the various efforts designed to adapt them to urban applications. More sprawl-specific studies are discussed in part two of the thesis. Also, much of the methodology detailed here, both in terms of land-use mechanics and traffic models, will re-appear in part three when the construction of sprawl models is discussed. These models fuse elements of both of these types of application, with emphasis on general land transition, but prompted by movement mechanisms.

Research in urban automata modeling is in early stages, and in some respects CA and MAS in urban studies still resemble their parents in the computing sciences. In the next chapter, we will introduce a framework for extending CA and MAS tools for *patently spatial* simulation. This framework—Geographic Automata Systems—retains the base functionality of general automata, CA, and MAS, but broadens the capabilities of those tools by means of uniquely spatial functionality. The GAS framework serves as the core for a range of simulation experiments designed to explore the geographic attributes of sprawl. A more in-depth discussion of sprawl is presented in part two of the thesis, but in the next chapter, the discussion of models and methodology continues where this chapter leaves off, with automata tools.

Chapter 5. Geographic Automata Systems

"And then the Vurt kicked in." (Noon 1993, p. 27)

Introduction

The advances detailed in Chapter 3 have supported a broad range of innovations in urban simulation, ushering in a new wave of models that remedy many of the weaknesses of conventional spatial and urban modeling methodology. Cellular automata and multi-agent systems are indicative of the state-of-the-art of this new wave, facilitating the development of models with unprecedented flexibility. As was discussed in Chapter 4, however, CA and MAS were developed, originally, as computing media. Their use in urban simulation requires a degree of adaptation. While geographers and other researchers engaged in urban studies have developed models tailored to urban applications, they are, for the most part, modifications to the general CA framework. MAS are comparatively ignored, save for traffic applications. Importantly, opportunities remain for developing spatial-specific models, with geographical foundations, from first principles. There is also a strong rationale for uniting CA and MAS approaches in a unified scheme, and for connecting automata approaches more directly with advances in Geographic Information Technologies.

This chapter introduces a spatially explicit framework for developing automata-based simulations, which was developed with Itzhak Benenson at Tel Aviv University (Benenson & Torrens 2003; Torrens & Benenson 2003). The framework offers significant flexibility for geographic applications, with innovations in relation to existing urban automata models. The chapter begins with a discussion of the limitations associated with the use of CA and MAS as urban modeling tools. A rationale for developing a new, geographically-founded, framework for urban automata models is then discussed. A proposed framework for Geographic Automata Systems is then introduced, with details about the specifications for the framework and its relationship to existing automata methodologies and potential urban applications. The fundamental components of the proposed Geographic Automata

Systems are discussed in turn: typologies, states and state transition rules, geo-referencing conventions and movement rules, neighborhood configurations and neighborhood rules. The chapter concludes with a discussion of the potential usefulness of the GAS framework, and its relationship to the advances discussed in Chapter 3. Having introduced urban modeling, the catalysts driving innovation in the field, and the state-of-the-art in terms of automata simulation, this chapter sets the stage for the rest of the thesis. Geographic Automata Systems are introduced as an innovative framework for urban modeling, before demonstration of the framework in relation to modeling suburban sprawl in North American cities in part three.

Limitations of cellular automata and multi-agent systems for urban simulation

Automata tools such as CA and MAS have numerous advantages for application in urban simulation. Compared to the conventional approaches discussed in Chapter 2, automata modeling is representative of a significant shift in the potency and usability of urban models. However, despite their suitability for simulating cities, there are limitations associated with CA and MAS, particularly when the tools are considered in isolation (Torrens 2000a; Torrens & O'Sullivan 2000b). These limitations are a consequence of an apparent disjunction between CA and MAS tools (Torrens 2003a), as specified for computing uses, and our theoretical and practical understanding of urban systems. In many instances, CA and MAS models of city systems do not behave as we would like them to, because the automata modeling framework does not facilitate specification of models that are fully representative of the systems being simulated. The research agenda for automata modeling of urban systems is in its infancy and many issues remain to be resolved (Benenson & Torrens 2004a; Torrens & O'Sullivan 2001). However, there are some specific limitations associated with the methodology of automata simulation that may not be readily solved with existing automata frameworks.

CA cells are not well-suited to supporting agency. Cells can be designed to mimic agency and agent-like behaviors, as was discussed in the context of traffic simulation in the last chapter. However, CA lack true agent-like functionality. Individual cells in cellular urban models are designed to articulate transition between states, based on the application of rules to information input to them from their neighborhood, e.g., 'initiate transition to a higher land value if the majority of cells in your neighborhood are of a higher value'. This is certainly intuitive—land prices do demonstrate propagation patterns. However, these types of rules ignore the real behaviors and processes governing phenomena like this in the real world. Real urban cells do not 'mutate' like cell cultures on a Petri dish (O'Sullivan & Torrens 2000). On the contrary, they change between states because of myriad interactions taking place between people and the urban fabric contained within those cells. Rules designed to mimic patterns do just that, but they are not representative of true system behavior. As we will discuss in part two of the thesis, behavior is essential to understanding phenomena such as sprawl.

The assumption of fixture for independent cells in a CA lattice creates other limitations. As discussed in Chapter 4, CA cells cannot move; CA can facilitate information exchange between cells—diffusion—but do not support the specification of true motion. The vehicle and pedestrian traffic models discussed in Chapter 4 mimic motion in a CA framework, by passing the presence of a "vehicle" through cell neighborhoods. This approach does actually provide close analogies with real-world traffic movement, when used with fine grain internal simulation clocks on the order of second-by-second time steps (Torrens 2004a). But, traffic models are an exception rather than the rule, and the methodology is inappropriate for modeling movement over large distances, or movement by displacement, for example in the case of migration or relocation. Again, this sort of long-distance movement is important for understanding sprawl.

The MAS approach is much more flexible and extensible than the CA approach. Nevertheless, it also has limitations in the context of urban simulation. Agents are not particularly well-suited to representing landscapes of spatial structures. CA are an excellent tool for encoding discrete spaces, in a variety of forms, into a model as a lattice of cells. There are few ways to generate such geometrical structures using agents, save to build a mosaic of static agents. As was evident in the case of the traffic models discussed in Chapter 4, agents that contribute to MAS must often be specified in an explicit environmental setting, and that environment is commonly represented distinctly and separately from agent entities.

Considered together, these limitations provide a rationale for designing new approaches for urban automata models. Yet, the significant advantages of the automata approach warrant that much of that methodology be retained.

A rationale for geographically-founded automata models

Automata tools have advantages for geographic simulation, but there has been relatively little exploration into the development of patently spatial automata tools. Geographers began to work with automata tools long after the approach permeated other disciplines. There was some early work in the 1970s, using automata in geographic models (Chapin & Weiss 1962, 1965, 1968; Nakajima 1977; Tobler 1970, 1979), but automata remained largely ignored in geography until the 1980s, when there was somewhat of a revival in interest (Couclelis 1985; Phipps 1989). However, it was not until the late-1990s that automata tools enjoyed popular use in geographical simulation (Batty et al. 1997; O'Sullivan & Torrens 2000). In recent years, there has been quite energetic use of CA tools in geographic research. MAS have a more recent history of use in geographical contexts, with just a handful of published models. CA and MAS have been applied to geographic problems, but there has been next to no research into the development of geographic automata.

Geographic Automata and Geographic Automata Systems

The framework that is proposed in this chapter starts with automata at its foundation, coupling general automata, CA-, and MAS-like approaches through spatial-specific functionality. The premise on which the framework is based is that of *objects* and their situation in *space*.

The general automata that were discussed in Chapter 4 were characterized by states, rules, and neighborhoods. We consider a new class of automata for spatial modeling—Geographic Automata (GA)—and consider simulations designed with many interacting GA as being specified as Geographic Automata Systems (GAS). GA differ from other automata in their characteristics and functionality; they are specified with explicit consideration of space and spatial behavior. GA retain the attributes of general automata, but the concept is extended with geography-specific characteristics:

- A typology of GA entities, based on spatial behavior, with a fundamental distinction between fixed and non-fixed GA
- A set of geo-referencing conventions that enable the situation of GA in space or in an environmental setting
- A set of neighborhood rules for expressing relationships between GA. Whereas traditional neighborhood filters are pre-defined and static, GAS rules are endowed with flexible rule-sets for determining agency and relativity, allowing those relationships to be varied in space and time
- A set of movement rules that consider the mobile functionality of the agent-based paradigm. These rules allow for the independent motion of GA in their simulated environments.

A GAS can be specified with several fundamental components:

$$G \sim (K\,; S, T_S; L, M_L; R, N_R) \qquad \text{Eq xxxviii}$$

G refers to a Geographic Automata System. K is a set of Geographic Automata types. S is a set of state variables, equivalent to the state variables of general automata. T_S refers to a set of rules that govern transition between states, S. L is a set of geo-referencing conventions, and M_L denotes a set of movement rules for Geographic Automata in the simulated system. R expresses the neighborhood specification for Geographic Automata, with N_R representing a set of neighborhood rules for determining neighborhood relationships, R.

These components operate dynamically in the context of a GAS simulation, with rules (T_S, M_L, N_R) determining transition between conditions in the state, geo-referencing, and neighborhood variables (S, L, R). Or, put another way, the state transition, movement, and neighborhood rules govern a Geographic Automaton's state, location, and neighbors at a given moment in time.

Types of Geographic Automata

GAS may be composed, at a fundamental level, of GA of different basic types (K). At an abstract level, we can make a distinction between *fixed* and *non-fixed* GA.

Fixed GA represent objects and entities that do not alter their location in the simulation over time. In this regard, fixed GA have analogies with CA cells. All manner of urban objects, for example, could be considered as fixed GA: schools, parks, railways, etc. Fixed GA are subject to transition rules as described in the previous section, except for movement rules (M_L).

Non-fixed GA represent entities that can shift location in space. Non-fixed GA are subject to all transition rules in a GAS, including movement rules (M_L). Also, the geo-referencing conventions (L) that characterize their location in space at a particular point in time will need to be updated dynamically, if a simulation evolves through time. Again, numerous urban objects can be represented using non-fixed GA: people, vehicles, transported goods, etc.

An urban system, or sub-system, will generally comprise objects of both fixed and non-fixed types. For example, a common retail "high street" may contain mobile entities in the form of pedestrians, cars, bicycles, etc. There may also be fixed objects that comprise the system: shops, sidewalks, crosswalks, etc.

Often, there are close relationships and co-dependencies between the characteristics of fixed and non-fixed GA; changes in the variables of either type of GA often instantiate changes in the other. A highway is an obvious example, where the velocity of moving vehicles (non-fixed GA) is governed, in large part, by the speed limit posted on signs along the individual links of that road, and the sequencing of traffic signals at junctions (signs and signals could both be specified as fixed GA).

States and state transition rules for Geographic Automata

State descriptors (S) for GA are equivalent to those used in general automata, CA, and MAS; they describe the conditions of the GA at a given point in time. States of geographical significance may be ascribed to GA, such as height, visibility, accessibility, etc. Similarly, type-specific (K) state variables may be introduced. In the context of non-fixed automata, these are likely to be associated with movement, e.g., inertia, velocity, direction, etc.

State transition rules (T_S) for GA operate like those used in general automata—as functions that govern shifts in the state variables of GA as a simulation evolves through time.

Geo-referencing conventions and movement rules for Geographic Automata

Geo-referencing conventions (L) in the GAS framework are specified as mechanisms for situating objects in space and storing spatial reference data as state variables within GA. The conventions will likely vary for different types (K) of GA. Geo-referencing is relatively simple for fixed GA; they can be referenced by their coordinate position, for example, and these values will persist over the course of a simulation run. Geo-referencing is less straightforward for non-fixed GA, however.

Fundamentally, we can distinguish between two means of geo-referencing for GA: *direct* and *indirect*. Direct geo-referencing refers to the situation of GA in a particular place at a particular moment in time—registering GA to the space that they *occupy* in a simulated environment. Direct location information may contain details such as point coordinates, polygon edge vertices, node identifiers, etc.

Indirect geo-referencing conventions refer to mechanisms for registering GA to *other objects* in a simulation—*pointers* that express the position of GA relative to other entities in a simulation. Indirect reference data could include variables such as, "visible from", "distance to", and "accessible from".

Land developers are a good example for illustrating direct and indirect geo-referencing. A land development firm, for example, may be referenced directly to its office address. Indirect conventions may also apply, relating that company's current development projects around the city to that fixed office.

Movement rules (M) are used in GAS to animate non-fixed GA in a simulation. Movement rules determine how GA navigate and are mobilized in simulated spaces.

These rules may be used to specify local-scale movement—locomotion such as walking, driving, flying, etc. These types of rules may focus on behaviors such as lane-changing, collision avoidance, and junction negotiation. Other movement rules can be designed to introduce larger-scale movement—migration, for example. Such rules may emphasize behaviors like push and pull factors, attraction, agglomeration, etc.

Neighbors and neighborhood rules for Geographic Automata

Neighbor relationships (R) in GAS are expressed using conventions from both CA and MAS approaches. Again, conventions may vary based on the types (K) of GA being related. Fixed GA remain static in space, and their neighborhoods can be defined using contiguous adjacency relationships, like the Moore and von Neumann neighborhoods associated with CA lattices in Chapter 4 (Figure 7). Other geometrical neighborhoods may be introduced, e.g., Voronoi polygons, Delauny triangulations, Archimedian tessellations, etc. Also, network-based neighborhoods may be introduced: links between nodes, edges between graph vertices, shortest paths between points, etc.

Neighborhoods may be dynamic for non-fixed GA, with configurations varying as a simulation run unfolds. A variety of variable neighborhoods configurations could be introduced: dynamic nearest neighbors, buffering distances between mobile entities, collision detection filters, etc.

Neighborhood rules (N_R) are introduced to determine transition between neighborhood configurations (R) over time. For the most part, neighborhood configurations for fixed GA will remain stable over the course of a simulation run. However, there are instances where they could change; the subdivision of land parcels is an obvious example. A variety of rules may be introduced to animate neighborhood configurations for non-fixed GA. Examples include market catchment rules for ice cream vans, rules determining visibility for pedestrian walkers, way-finding rules for shoppers in a supermarket.

Conclusions

A new framework for building geographical simulations with automata tools has been introduced in this chapter—Geographic Automata Systems. The GAS framework is based on a general automata foundation, and fuses elements of the CA and MAS approaches discussed in Chapter 4, but adds additional—uniquely spatial—functionality. The framework allows for the specification of models composed of independent Geographical Automata, coupled as GAS to perform simulation tasks.

GA retain features of general automata: state variables and state transition rules; neighborhoods; and an evolving temporal environment. Borrowing from agent-based approaches, these features are interpreted in a flexible context; this is a departure from the comparatively strict and sometimes static specification of popular CA approaches (Torrens & O'Sullivan 2001). Additional, spatially explicit, operations and mechanisms are introduced. First, GA are distinguished by geographical classifiers, based on fundamental spatial attributes—whether they support movement or not. While this seems intuitively obvious, it is nonetheless a distinction that is absent from much of the methodology currently used in automata-based urban simulation. Second, stemming from the elucidation of movement, geo-referencing conventions account for the absolute or relative location of GA in a GAS, at a particular place or in relation to other model entities, in a given time-step in a simulation's evolution. Third, specific movement rules are also introduced for articulating the motion of GA on local scales and over large distances. Fourth, neighborhood rules are featured, enabling neighborhood structures and relationships to be reconfigured, dynamically, as a simulation evolves.

The GAS framework outlined in this chapter offers significant innovation over existing automata-based methodologies that are popularly employed in the development of urban models. In particular, it addresses several of the flaws of cellular urban models, as mentioned in the literature in the field (Torrens & O'Sullivan 2001). Importantly, it infuses a patently geographical toolset into the methodology. Rather than relying on affinities between automata and spatial (geometrical) structures, geography is placed at the very heart of the methodology—geography becomes the fundamental foundation on which simulations are based. Furthermore, the GAS framework provides an intuitive and spatial basis for unifying

CA and MAS methodologies in a symbiotic fashion; it enables the expression of environment-environment, agent-agent, agent-environment, and environment-agent dynamics and interactions.

The GAS framework is also compatible with current developments in fields peripheral to urban simulation, and capitalizes on the developments mentioned in Chapter 3. Ideas from complexity studies can be used to determine how independent GA interact to form GAS through processes such as self-organization, phase shifts, and positive feedback. Object-oriented paradigms serve as an excellent vocabulary for expressing GAS components in software code, as do entity-relationship approaches. In many ways, GAS can be regarded as extending Geographic Information Science, uniting spatial modeling and spatial analysis (Benenson & Torrens 2003; Torrens & Benenson 2003). In many cases, relationships between spatial analysis and modeling are loosely-coupled—software packages are used cooperatively, for example. GAS can be utilized in that manner; GIS can be used to store geo-referencing information. However, GAS also offer a framework for tightly-coupled relationships between modeling and analysis. The spatial rule-sets introduced in GAS can be expressed using familiar methodology from Geographic Information Science, weaved directly into the simulation methodology itself.

Elsewhere, the relationship and suitability of the GAS framework to other popular automata models has been examined (Benenson & Torrens 2003; Torrens & Benenson 2003). In that work, it was found that *all* urban automata models can be expressed as GAS. GAS is also used as the framework for designing re-usable simulation software in that work.

Part three of the thesis details the application of the GAS framework to urban simulation, modeling the mechanics of suburban sprawl in the context of North American cities, from a variety of perspectives and at varying scales of observation. The purpose of those sections of the thesis is to demonstrate how GAS work in an applied context, where GAS offer innovation over existing modeling methodologies and technologies, and to consider how GAS might be used as the basis for experimenting with ideas and hypotheses about complex contemporary urban trends and problems. It is to those sorts of contemporary urban problems that attention now turns; part two of the thesis focuses on the phenomenon of suburban sprawl in American cities.

Part two: substantive material

Chapter 6. Sprawl characteristics, consequences, and causes

"But here the mirrored ziggurats down Lázaro Cárdenas flow with the luminous flesh of giants, shunting out the night's barrage of dreams to the waiting *avenidas*—business as usual, world without end." (Gibson 1993, p. 1)

Introduction

Part one of the thesis provided an overview of models and methodologies for spatial and urban simulation. Chapter 5 introduced a new, proposed, framework for simulating cities. This framework offers potential for overcoming many of the traditional flaws of urban models, and for remedying some of the limitations of automata-based tools. The framework—Geographic Automata Systems—is patently spatial in design, offering unique and innovative functionality for modeling geographic phenomena. In part three of the thesis, the framework will be applied to the simulation of suburban sprawl in North American cities.

Before developing an environment that simulates suburban sprawl, as an example of contemporary urban dynamics, it is first necessary to understand the phenomenon on a conceptual level. Part two of the thesis focuses on substantive issues relating to sprawl. This chapter presents a review of the literature pertaining to sprawl in North American cities, with particular attention to its characteristics, consequences, and causes. The next chapter looks at models of sprawl and sprawl properties. Sprawl is not unique to North America; it has been studied in European contexts for a long time (Gottman & Harper 1967). However, its characterization and manifestation in North America is much lower in density and scattered in nature than in Europe (where it is generally referred to as peri-urbanization), and is perhaps more significant in terms of its costs and associated problems.

Sprawl is commonly characterized in a very simplistic way—descriptively and textually—as voracious expansion of the urban periphery into previously undeveloped agricultural areas or natural resource land, at uniform low densities and in a fragmented pattern. Sprawl is one of the most important urban issues facing contemporary cities. This is particularly true in North America, where urbanization is a relatively recent phenomenon in many areas. Sprawl is sweeping through North America at unprecedented rates. It even featured as a campaign issue in the 2000 presidential elections. However, there is little empirical evidence to characterize the attributes of sprawl—it is not quite obvious how to identify it. By extension, this makes it quite difficult to develop sprawl models. Nevertheless, some characteristics can be attributed to sprawl, although the literature base for research in the area is very broad.

The remaining sections of this chapter will explore the characteristics, consequences, and causes of sprawl, as a prelude to specifying models of suburban sprawl in part three of the thesis. Characteristics of sprawl are discussed with reference to density, scatter and fragmentation, open space, aesthetics, accessibility, and dynamics. The discussion of characteristics prefaces selection of model variables and the identification of model outputs in later chapters. The literature about the consequences of sprawl is also reviewed, with attention to some of the potential benefits of sprawl, as well as associated costs. Direct costs are considered, including pubic infrastructure costs. Externalities are also considered, including air pollution, downtown decline, loss of land, adverse effects on water quality, and reduction in quality of life. Relatively little research has been performed on the causes of sprawl in North America, simply because it has become a recent concern, and because of its complexity. Nevertheless, some idea of the mechanisms driving suburban sprawl is necessary if rule-based sprawl models are to be devised. Likely causes of sprawl are explored: population growth, consumer preferences, decentralization trends in economic activity, transport and telecommunications influences, development, planning, public policy, and political fragmentation. The chapter draws to a close with a summary and some conclusions.

Characteristics

Put simply, North American sprawl manifests itself as very rapid urban growth on the edge of metropolitan areas, with characteristic low density and scattered spatial distribution. This is a very cursory definition, of course. Theoretical debates about sprawl in North America have generated a wealth of discussion around the issue of characterization. Echoing Ewing's (1997) comments, however, we still lack a *working* definition. Sprawl is a practical, real world concern. Arguably, there is a sense of urgency attached to the sprawl problem. Sprawl is almost universally associated with disdain in the United States, where city planning agencies and citizen advocacy groups are scrambling to control its spread. A variety of characteristics have been proposed in the literature, but for the most part they are descriptive rather than quantitative. This makes it difficult to plan measures to contain sprawl and develop policies to mitigate its spread. It also makes it difficult to determine the variables that might feature in a model of sprawl. Nevertheless, looking across a broad literature base, some empirical characteristics can be identified. Several of those will feature in the GAS models of sprawl described in part three of the thesis.

Density characteristics

The most frequently cited characteristic of North American sprawl is low-density development. However, density is among the fuzziest and most speculated about definitions of sprawl. In particular, a number of important considerations are unclear in the current body of literature devoted to sprawl: the best variable to use in representing density, at what value of density a city might be regarded as sprawling, the scale at which density should be measured, and the extent of the space over which density might be characterized.

A number of variables have been used to represent density, most commonly density of housing units, population, and employment. The Lower Mainland Regional Planning Board of British Columbia, for example, characterized low-density sprawl in Canada as urban areas with *population densities* of 0.3 to 0.5 people per acre. At these densities, they argued, parts of the city are less than adequate for efficient service provision and too high for true agricultural development (Ledermann 1967). On the other hand, in their work, *The Costs of Sprawl* (1974), the Real Estate Research Corporation (RERC) referred to low-density sprawl as a *housing density* of 1,360 units per square mile (Real Estate Research Corporation 1974).

The scale at which density is studied is equally important in any consideration of a density measurement. Depending on the scale of observation—the metropolitan area, a district within the city, a neighborhood, parcel-by-parcel—the value reported for density may vary considerably.

Another topic of contention in sprawl literature is the extent over which densities should be measured (Gordon & Richardson 1997a). Should an analyst use the total area of a city in her density calculation (gross density), or should she omit areas upon which people could not possibly reside such as open water, parks, wetlands, cemeteries, industrialized areas, disposal sites, etc. (net density)? While calculating net densities would seem intuitively superior to gross densities, an analyst must consider that the exclusion of areas such as open water and industrial areas—which have a direct impact on housing costs—might bias measurements of density. However, Zielinksi (1979) has argued that such biasing may be unavoidable and favors using gross density as a measure because it is likely that the negative and positive affects of omittable land uses would most likely balance on the whole.

Aside from these issues, there is also the problem of looking at density in abstract terms; this detracts from the geographical element of the variable. It may be more appropriate to study how the density *gradient* of a city has changed over time. If a city has sprawled, its gradient will exhibit a marked flattening from one period to the next. Some more localized measures of density, such as density within a specified radial distance of a given location, may also be appropriate sprawl indicators.

Scattered characteristics

The scattered characteristics of North American sprawl manifest themselves in a variety of guises: fragmentation, leapfrogging, discontinuous development, dispersal, and piecemeal development; sprawl is an inherently spatial problem. Essentially, these characterizations amount to the same thing—tracts of developed land that sit in isolation from other undeveloped tracts (Lessinger 1962). The scattered nature of sprawl can be both costly and unsustainable. Because scatter isolates residences and opportunities, travel times in sprawled areas grow, as do associated environmental damage and energy consumption. Also, the cost of providing essential urban services and infrastructure—wastewater facilities, water pipes, telecommunications networks, garbage collection, emergency services, roads, schools, etc.—in scattered areas is much greater than would be the case in more compact neighborhoods (Ewing 1994).

Other authors assert that great capital expenditures must be pumped into the provision of urban services in sprawled areas even at the initial time of development, especially when much of the land may be left vacant (Harvey & Clark 1965).

Differentiating scatter from economically efficient "discontinuous development" (Ewing 1994) can be difficult, however. The distinction involves weighing up several components of scatter: the quantity of land bypassed in the initial development wave, the length of time that land is actually withheld from development, and the ultimate use of the land (Ewing 1997). The temporal components of scatter are another consideration. Sprawl is a dynamic phenomenon; what looks like sprawling suburb today could well evolve into compact and sustainable development in later years as the pace of urban extension drives developers to fill-in previously undeveloped sites (Peiser 1989).

Another way to consider the scattered nature of sprawl is via its *fractal dimension.* This can help us to characterize the space-filling capacities of urban development. If we consider possible integer dimensions for an urban area, we can distinguish a city with a dimension of zero (a city existing on a point). A city with a dimension of one (a city existing as a line) might correspond to something like Soria Y Mata's 'La Cuidad Lineal' or Frank Lloyd Wright's 'Mile-High Skyscraper City'. A city with a dimension of two (fully occupying a two-dimensional plane) might also be conceived, such as Wright's 'Broadacre City'. Indeed, we might imagine a city that fully occupies three dimensions, such as Dantzig and Saaty's 'Compact City' (Batty & Kwang 1992; Batty & Longley 1994). The existence of one-, two-, and three-dimensional cities might easily be conceptualized in a theoretical context, but in a practical setting such forms are unlikely to exist. Most cities do not occupy a single or multiple dimensions fully. On the contrary, cities usually occupy fractions of a dimension or exist at some stage between multiple dimensions. Fractal dimensions are dimensions that operate *between* integer dimensions. In this sense, fractals are a good way to characterize the scattering of development, which would take on fractional values of dimension (as opposed to integer values) in such a scheme. (Indeed, we will use fractal dimension in Chapter 8 to evaluate the extent of sprawl in our simulated cities.)

Open space characteristics

The terms sprawl and suburbanization are often synonymous but sprawl is commonly characterized as a *wasteful* form of suburbanization (Ewing 1997). Sprawl lacks functionality. The term "functional" in this context implies some useful public service for an urbanized space (Ewing 1994). Sprawl falls short as a functional use of space for a variety of reasons. Ribbon sprawl creates walls of commercial activity that restrict access to much of the open space around them (more about this later). With low-density sprawl there are provisions of open space, but the space may not be functional because it is in private hands; essentially much of it serves no public use. Leapfrog development is equally deficient in this regard. Scattered urbanization leaves lots of open space in the gaps, but there is little functionality, again because it is held in private hands. The space cannot even be used for agricultural purposes because it is generally worth too much money to be used as farmland (Ewing 1994).

Aesthetic characteristics

The aesthetic characteristics of sprawl—images that it conjures in the mind's eye—are one of the less tangible qualities of the phenomenon, yet they are also amongst its key components. In North America, urban sprawl is widely regarded as a lazy and undisciplined expression of urbanization (Gordon & Richardson 1997a), almost universally met with criticism and distaste:

> "Urban sprawl, roller-painted across the countryside is often without form, grace, or a sense of community. Planning philosophies aimed to strike down this amorphous creature should only gladden our hearts" (Lessinger 1962, p.159).

> "It is not just the sentimental attachment to an old sledding hill that has you upset. It is the expectation, based upon decades of experience, that what will be built here you will detest. It will be sprawl: cookie-cutter houses, wide, treeless, sidewalk-free roadways, mindlessly curving cul-de-sacs, a streetscape of garage doors—a beige vinyl parody of *Leave it to Beaver*. Or, worse yet, a pretentious slew of McMansions, complete with

the obligatory gatehouse. You will not be welcome there, not that you would ever have reason to visit its monotonous moonscape. Meanwhile, more cars will worsen your congested commute. The future residents will come in search of their American Dream, and in doing so will compromise yours." (Duany *et al.* 2000, p. x)

One of the often-vocalized criticisms of sprawl, apart from a widespread dislike of low density and scatter, is "ribbon sprawl," generic fast-food–lined alleys of car parks and neon signs that burn phosphorously, long into the night. The phenomena of ribbon sprawl, or "retailscape" (Gordon & Richardson 1997a) is closely related to scatter, although it is radial rather than planar in form. Ribbon sprawl generally manifests itself as strips of commercial development (normally retail outlets and related premises) that flank the sides of highways and main thoroughfares. Exit-parasitic retail development is an associated component: the clustering of retail establishments (hotels, gas stations, fast food restaurants, etc.) close to highway exit ramps (Torrens 1998). Ribbon sprawl is composed of segments of developed land that are compact in themselves but which extend axially and leave the intervening space undeveloped (Harvey & Clark 1965). This creates walls of commercial development (often buffered by large 'seas' of car parking) that restrict access to much of the space around them.

Accessibility characteristics

Sprawl is also commonly characterized in terms of accessibility. Sprawl-type urbanization has poor *residential accessibility* because residents are distanced from out-of-home activities. At the same time, sprawl is characterized by poor *destination accessibility:* out-of-home activities themselves are far from each other (Ewing 1997). Why is sprawl so deficient in terms of accessibility? The scattered nature of sprawl is a contributing factor. Residents must pass undeveloped tracts of land in order to navigate sprawling areas. Retailscape is another factor: along ribbon developments, motorists must traverse a plurality of linearly arranged commercial uses (usually on crowded arterials) on the way from one shop to the next—the opposite of one-stop shopping (Ewing 1997). Of course, density also contributes to accessibility problems. For low-density development, everything is dispersed, making all trips longer. The average suburban household in the United States drives 30% more on a yearly basis than a central city counterpart would. In the United States at present rates of average Vehicle Miles Traveled (VMT), this translates to 3,300 more miles per year at an added cost of $753 per year per household. Suburban residents also spend 110 more hours per year driving cars than central city residents do, on average; the equivalent of three weeks work (HUD 1999).

Dynamic characteristics of sprawl

Sprawl is commonly treated as a static phenomenon. This is inaccurate; sprawling areas of American cities are at the forefront of dynamic urban growth. By misinterpreting sprawl as static, planners and policy makers risk making incorrect judgments.

> "The sprawl of the 1950s is frequently the greatly admired compact urban area of the early 1960s. An important question on sprawl may be, "How long is required for compaction?" as opposed to whether or not compaction occurs at all...The concept of time span is important in the identification and measurement of sprawl. The application of static measures to dynamic areas can easily result in the misidentification of an area as sprawl when it is really a viable, expanding, compacting portion of the city" (Harvey & Clark 1965, p.6).

To understand sprawl as a dynamic phenomenon, it is necessary to understand change in each of the characteristics discussed thus far—how they evolve through time and through interaction, how their dynamics operate across scales, and how their spatial distribution varies.

The consequences of sprawl

The costs of sprawl in North America have been debated wildly in the literature (see (Ewing 1997; Gordon & Richardson 1997a) for the classic 'sprawl show-down'). Like all good empiricists, I would prefer to remain objective on the topic and wish to dodge any questions of whether sprawl is 'good' or 'bad'. However, it is necessary to discuss the topic nonetheless. Although, from a modeling standpoint it is perhaps best to remain emotionally detached from the phenomena being simulated, it is important that any urban simulation be placed in its policy context. To that end, this section provides a review of the literature on the costs of sprawl. It is worth mentioning that there is little in the way of hard evidence to support arguments about the consequences—or 'costs'—of sprawl, largely because there is so much room for debate regarding the actual characteristics of sprawl. Predictably enough, the section is more verbose on the subject of disadvantages, simply because more has been written on the topic. Nevertheless, I have made an effort to be even-handed in my treatment.

The benefits of sprawl

Most studies are savagely critical of sprawl, yet it is important to face facts: there are many advantages to sprawl-like development, both for local residents and society as a whole. There are a number of papers that argue in favor of sprawl, particularly those from researchers at the University of Southern California (Gordon & Richardson 1997a, b; Peiser 1989). While some would contest the arguments offered in those papers as being sensationalist—"Peter Gordon and Harry Richardson (G & R) have made a cottage industry out of challenging, time and again, planners' steadfast belief in compact development" (Ewing 1997, p.107)—they do have a point: sprawl is not all bad. In a general sense, once a city grows beyond a certain size, a polycentric urban form may be more efficient than compact and more centralized development (OTA 1995). (The findings in Chapter 8 support those sentiments to some extent.) There are a number of reasons why this may be the case, particularly the clustering of land-uses and associated reduction in trip lengths and congestion.

For individual city-dwellers, sprawl has many advantages. Housing is often more affordable at fringe sites in the city, largely because land costs are cheaper at peripheral locations. Sprawling locations can be regarded as offering nicer living environments; as we will see later, they are often perceived to be less polluted and congested in the United States, while being free from crime and social tensions.

Ironically, trends toward automobile dependence in sprawling areas have been cited as beneficial, both for individuals and society. By catering to an automobile-dominated society, sprawling parts of the city offer several advantages to the individual car owner: comfort, flexibility, low door-to-door travel time, freight carrying capacity for shopping trips, cheap long-distance travel, and the aesthetic benefits of separated land-uses. Equally, there are a number of broader benefits for society. Sprawl areas may offer more options for businesses seeking new locations, thereby increasing economic efficiency. Consumer access to 'big box' stores may also increase economies of scale in retailing. Also, non-transit automobile commutes to work releases employees from dependence on transit timetables, thereby allowing for more flexibility in work schedules and an associated increase in the efficient use of human capital (OTA 1995).

The costs of development in sprawling areas may be lower than those for comparable sites in central cities because of the relative expense of rights of way and disruption costs (OTA 1995). Other authors have also argued that scattered development may actually promote higher development densities in some cases (Peiser 1989). Parcels of land that have been skipped by the development process have the potential to rise in value as sprawled areas of the city are subjected to infill development over time, and as such would be more likely to be developed at higher densities. Some authors even contend that sprawl offers environmental benefits. They argue that sprawled parts of the city are better at dispersing air pollution because they are spread over large areas (Bae & Richardson 1994). This is very plausible, particularly in the area in which that research was conducted: the Los Angeles Basin.

The disadvantages of sprawl

Direct costs

The literature regarding the negative consequences of American sprawl is a much more substantial body of work, at least in terms of volume, than that arguing the merits of the phenomenon. Of course, it is easy in an intuitive sense to think of reasons why sprawl might be costly. However, there remains little in the way of hard evidence to say—definitively—whether sprawl is more expensive than more compact forms of development. There are some convincing theoretical arguments that support this contention though. Most of the arguments about the expense of sprawl focus on questions of whether it costs more than more compact forms of urban development, for example, that most often seen in central cities, or perhaps 'New Urbanist' forms of urban design (Duany *et al.* 2001; Katz 1993), and queries regarding who pays for those costs.

Studies suggests two principal categories and some further classifications that we might use to group the costs of sprawl (OTA 1995) (Figure 17): principally, direct costs and externalities. Direct costs include the costs of providing pubic infrastructure and services. Externalities include loss of land to development, air pollution, downtown decline, degradation of the quality of life, affects on water quality, and ecological implications.

Public infrastructure costs

The costs of providing essential public services to sprawling communities could be regarded as both a direct neighborhood and community cost. Generally, it is considered that reduced densities incur more costly service provision—roads, wastewater facilities, education, police and fire, garbage collection, etc.—because the buildings that need to be supplied are further away from central nodes of facility provision and also from each other.

Research into the direct service costs of sprawl is relatively widespread, possibly because it is among the easier of the costs to measure empirically. A number of studies have been performed in the United States; generally they evaluate the relative costs of urban development at various densities, or the costs of certain spatial patterns for development (American Farmland Trust 1995; Frank 1989; James Duncan & Associates *et al.* 1989; Real Estate Research Corporation 1974). The results of those works are summarized quite thoroughly in other publications (Benfield *et al.* 1999; OTA 1995). Generally, the research concludes that compact and contiguous development is more cost-efficient than low-density, scattered development.

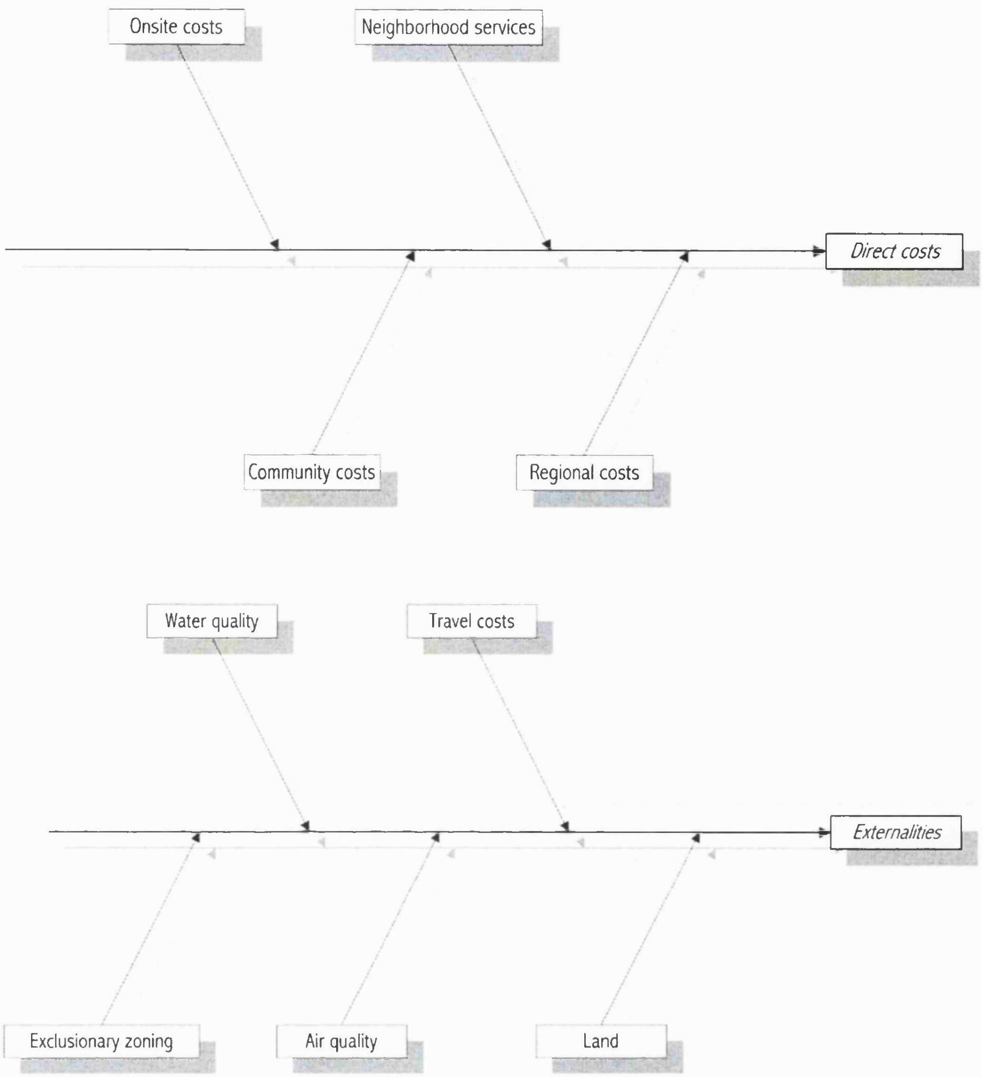


Figure 17. The costs of sprawl

Of course, even with empirical evidence, there remains contention as to the appropriateness of these findings. There is some evidence that single-family homes are more expensive to service when compared with more compact development types (Figure 18). However, this only becomes a problem if the development fails to pay for itself, i.e., if the costs of service provision are not directly passed on to the owner of the building or indirectly absorbed through the developer. However, the argument that costs will be endured by the neighborhood and/or community through the *maintenance* of these facilities is convincing enough (OTA 1995).

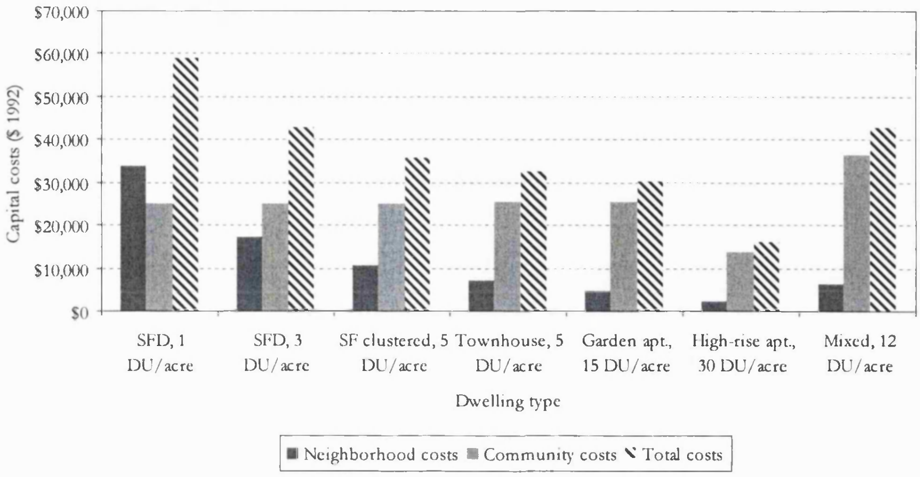


Figure 18. Costs of providing capital facilities[1]

(Source: Environmental Protection Agency 1993.)

A major issue of contention and source for debate is the question of just exactly who pays for urban services in sprawled areas of the city. The cost of services to the consumer such as electricity, gas, water, and cable television are all billed at rates that are independent of distance from central facilities. There is little, if any, spatial differentiation in the pricing of utilities. The only distinction is between residential and commercial users (Guy *et al.* 2001). Because of this, residents in central cities could be regarded as subsidizing the facilities of suburban residents. At the community level, work has demonstrated that sprawl-like development rarely "pays for itself" (James Duncan & Associates *et al.* 1989). Cost-revenue studies in Florida found that scattered and linear developments had much lower ratios of cost-to-revenue when compared with relatively compact and contiguous development. Also, in some cases, where suburban jurisdictions are within the boundary of central jurisdictions, a central city may end up indirectly subsidizing sprawl. This happens particularly in Western states, where annexation of suburban territory has been common.

[1] SFD: single family dwelling; SF: single family; DU: dwelling unit.

However, yet again the argument swings both ways: in the United States, subsidy of this kind is not likely to occur in Eastern states, where fragmentation is uncommon (OTA 1995). Also, a careful examination of the results from other studies reveals that virtually *all* development, including downtown, fails to pay for itself (James Duncan & Associates *et al.* 1989). Anyway, it is difficult to discern where subsidies are actually coming from, even when they have been found to exist. Many of the costs of service and infrastructure provision are covered by other local governments or other levels of government (e.g., state and Federal) (OTA 1995).

Externalities

Air pollution

Sprawl has been blamed for negatively influencing urban air quality. As we will examine later when we come to discuss causes, sprawl is associated with increased travel, particularly via automobiles. The scattered and fragmented nature of sprawl renders distances and the need for travel greater between locations, with associated influences on vehicle emissions and energy consumption.

Vehicles are known to release pollutants to the atmosphere as a by-product of organic fuel combustion. The average car releases 8,800 pounds of CO_2 to the air per year in the United States, and the figure for 'light trucks' (now the most popular class of new vehicle purchased in the United States) is almost twice that value (Benfield *et al.* 1999). Vehicles emit 62% of U.S. carbon monoxide, 26% of volatile organic compounds (VOC), 32% of nitrogen oxides (each of which have been linked to ozone smog), and 50% of benzene and formaldehyde (both linked to cancer) (Benfield *et al.* 1999). Although emissions have remained stable over time and in some instances have exhibited a marked reduction (Figure 19, Figure 20), vehicle-related pollution is still costly; a 1998 EPA report estimated the costs of traffic-related particulate pollution to be $20 billion to $64 billion per year (Benfield *et al.* 1999, p.58).

However, others urge caution in linking transport and air quality, arguing that the relationship between the two is far from clear (OTA 1995). Additionally, it has been argued that the longer trip distances generally characteristic of sprawled areas, which facilitate faster speeds, may be more efficient in fuel consumption (Bae & Richardson

1994). Moreover, air sheds in sprawled areas may be better at dispersing pollutants, simply because they are scattered over a wider and less built-up area.

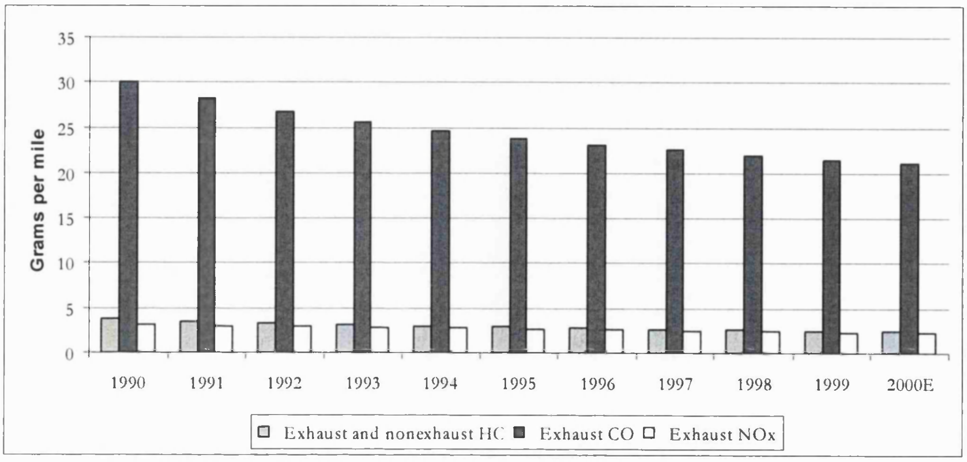
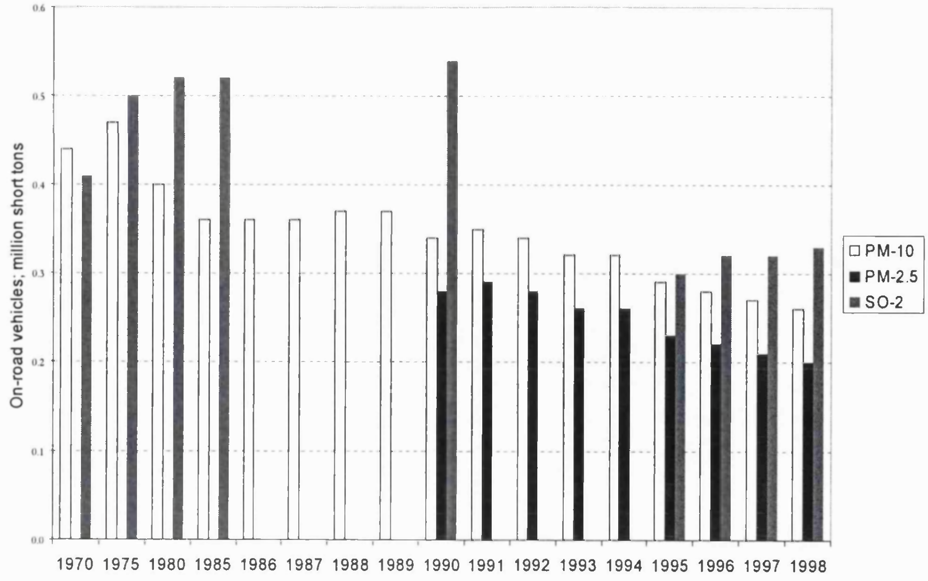


Figure 19. Estimated National Average Vehicle Emissions Rates[2]

(Source: U.S. Environmental Protection Agency, National Vehicle and Fuel Emissions Laboratory.)

[2] HC: hydro-carbon; CO: carbon monoxide; NOx.: nitrous oxides.



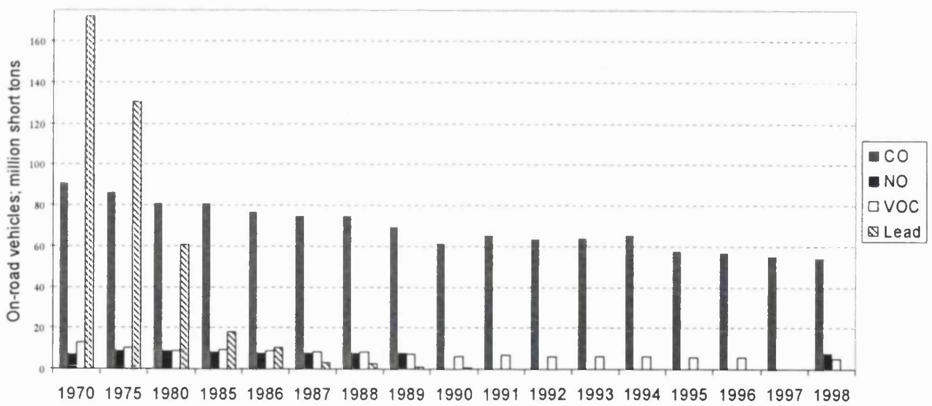


Figure 20. Estimated U.S. national average vehicle emissions[3]

(Source: Environmental Protection Agency 2000.)

[3] PM-10: particulate matter less than 10 μ in size; PM-2.5: particulate matter less than 2.5 μ in size; SO-2: sulfur dioxide; CO: carbon monoxide; NO: nitrous oxide; VOC: volatile organic compounds; where values are zero, data was unavailable. Values in million short tones, except lead, which is in thousand short tons.

Downtown decline

It is clear that for many metropolitan areas in North America, the sprawling periphery appears to be flourishing while central city counterparts are left languishing (Table 1). There may be justification for laying the blame for central city decline, or at least for exacerbating existing difficulties, with sprawling suburbs.

The spatial mismatch hypothesis is one justification (Hodge 1986). It refers to separation of jobs from potential employers. Jobs are often located in greater abundance in the suburbs, while potential employees remain in the central city, with relatively little means to reach work.

> "Unfortunately, most candidates for moving from welfare to work do not live close to available jobs. Instead, three-quarters of welfare recipients live either in center cities or rural areas, with urban poverty growing at the most rapid rate... however, job opportunities are now overwhelmingly located in suburban areas. In some regions (Chicago, Cleveland, Dayton, Detroit, Greensboro, Louisville), suburban jobs growth accounted for 100 percent of overall metropolitan job growth during the first half of the 1980s." (OTA 1995, p.125) (Statistics quoted from Hughes & Sternberg 1992)

There is an irony in the spatial mismatch of jobs and workers: many of the policy treatments devised to remedy the imbalance focus on supplying public transit facilities to link central areas with suburbs (with particular emphasis on transit flows from downtown to peripheral locations at peak hour congestion times). However, "suburban jobs are increasingly located in areas that lack the population and activity densities that justify transit routes" (OTA 1995).

Table 1. Metropolitan joblessness (Source: HUD 1999.)

	City unemployment					Suburb unemployment				
City	**1970**	**1980**	**1990**	**1992**	**1998**	**1970**	**1980**	**1990**	**1992**	**1998**
Atlanta	3.9%	8%	7.6%	10%	5.2%	2.6%	4.2%	4.7%	6%	3.1%
Buffalo	6%	13.1%	8.6%	12.2%	8.7%	3.9%	8%	3.7%	5.5%	3.8%
Chicago	4.4%	9.8%	8.4%	9.5%	5.6%	2.5%	4.8%	4.7%	6.2%	3.5%
Los Angeles	6.9%	6.8%	6.7%	11.1%	7.3%	5.6%	5.5%	5.4%	8.9%	6.1%
Miami	4.3%	6.1%	11.2%	15%	9.6%	3.3%	4.5%	7%	9.4%	5.7%
Milwaukee	4.1%	6.9%	5.7%	6.3%	4.8%	2.8%	4.2%	3.3%	3.8%	2.4%
New York	4.2%	7.7%	6.9%	11%	8%	2.6%	4.4%	3.4%	6.3%	3.9%
Philadelphia	4.6%	11.4%	6.3%	8.9%	6%	2.8%	5.7%	4.2%	6.7%	3.6%
Phoenix	3.8%	5.5%	4.9%	7.2%	2.9%	4.5%	5.9%	4.8%	6.9%	3.1%
Portland, OR	3.8%	6.3%	4.3%	6.3%	2.8%	2.7%	5.3%	3.5%	5.1%	2.1%
Seattle	8.2%	5.8%	4.1%	7.5%	3.5%	8.1%	5.6%	3.3%	6%	2.9%
St. Louis	6.4%	11.1%	8.4%	8.3%	7.2%	4.1%	6.6%	5.4%	5.4%	3.7%
Washington, D.C.	3.7%	6.6%	6.6%	8.6%	8.6%	2.1%	3.6%	2.7%	4.8%	2.6%
Top 10	4.7%	7.8%	7%	10%	6.2%	3.8%	5.6%	4.8%	6.9%	3.8%
Top 50	4.7%	7.2%	6.3%	8.7%	5.4%	3.7%	5.2%	4.4%	6.2%	3.5%
All (329) MSAs	**4.6%**	**7.1%**	**6.3%**	**8.5%**	**5.1%**	**3.8%**	**5.6%**	**4.7%**	**6.6%**	**3.7%**

Loss of land

The occurrence of sprawl on the urban periphery is particularly problematic in North America. In many cases, expansion on the urban fringe comes at the expense of agricultural and resource lands. A 1995 study found that from 1982 to 1992, an average of 400,000 acres of prime farmland (defined as land with the best soils and climate for growing crops) were lost to suburbanization in the United States (American Farmland Trust 1995). Additionally, 26,600 acres of unique farmland (defined as land used for growing rare and specialty crops) were lost each year over the same period. "Put another way, for each acre of prime or unique farmland that is being saved by various farmland protection programs across the count[r]y, three acres are lost to development." (Benfield *et al.* 1999) The problem of agricultural encroachment by urban development is compounded by the coincidence that many of the areas that are best suited to growing crops are also well suited to 'growing houses' (Ewing 1994). As prime land is converted to urban uses, land that is less viable for agricultural purposes is being brought into production, with the result that forests and wetlands are often lost. Once again, however, those claims are being countered in some instances. Some authors contend that U.S. agricultural surpluses negate any need for concern regarding the loss of agricultural land to development (Gordon & Richardson 1997a). Losses in terms of the *amenity* value of agricultural land are less easy to dismiss, however.

The loss of habitat land to sprawl is a related problem. Development in former forested areas, wetlands, or resource lands can, in some instances, disrupt ecosystems with negative impacts on flora and fauna in those areas. Rare and endangered species are particularly sensitive to shocks in the ecosystem. The scattered nature of peripheral sprawl is of particular concern because it leaves behind isolated patches of habitat that are often only suitable for generalist species that are already in abundance (Ewing 1994).

Water quality

A further externality commonly associated with sprawl relates to its impacts on water quality. As we have already seen, peripheral urbanization is often responsible for converting land with agricultural or resource land covers to urban uses. In some cases, this can exacerbate an area's susceptibility to non-point source pollution. Non-point source pollution occurs when water passes over a surface or through the ground; in the process pollutants and other deposits may be swept along and introduced into the groundwater. The pollutants that are collected may include sediment, pathogens, nutrients such as nitrogen and phosphorus, heavy metals, pesticides, and non-degradable debris (Benfield *et al.* 1999). Non-point source pollution can have several negative impacts upon water quality. It has the potential to damage the aquatic fauna in streams, as well as polluting soils and harming vegetation. Hydrological systems are also in danger from thermal pollution. This occurs when the temperature of water running over a surface is higher than that into which it flows and associated damage to aquatic ecosystems takes place.

Sprawl may also affect water quality. The major cause of non-source point pollution is impervious surface: artificial surfaces such as concrete, asphalt, and Tarmac that provide lower porosity than organic surfaces such as soil, vegetation canopies, wetlands, and grasslands. Sprawl manifests as a conversion from formerly organic surfaces to these artificial forms, often in areas closest to natural resources. However, it is likely that more dense and compact forms of development are more harmful to the environment, at least in terms of water quality. The area of impervious surface in densely urbanized areas is generally greater in extent compared to that in sprawling suburbs. Also, because sprawl leaves isolated tracts of land undeveloped, the range of impervious surfaces is punctuated with organically surfaced patches. Also, the statistics used to measure impervious surface—total impervious surface—may be inappropriate for use in sprawling areas. Effective impervious surface—impervious surface that is connected directly to streams naturally or through roads—may be better as a measure. Connectivity of impervious surface is related; it is measured as median distance of the impervious surface in a basin to the closest road or stream through the flow-path (using flow direction calculated through a digital elevation model) (Alberti 2001).

Quality of life

It is worth briefly mentioning that, in addition to the other costs of sprawl that we have mentioned this far, there exist several speculative arguments about the deterioration of quality of life in sprawling suburbs. The standard dispute is that sprawl promotes isolation, results in a loss of community, and fosters a lack of civic engagement (Benfield *et al.* 1999). While this may well be the case, there exists little empirical evidence to confirm or refute these claims, and the intangible nature of many of these arguments make them difficult to quantify.

Causes

In the previous sections we have explored various characteristics for recognizing and describing sprawl as well as various consequences associated with the phenomenon. Some of these characteristics will feature as variables and outputs in the sprawl models discussed in part three of the thesis. Now we turn to some of the mechanisms that have been proposed as engines of sprawl: we will attempt, at least in a conceptual sense, to tie *causes* to characteristics and costs. Many of these will be modeled as geographic mechanisms in the simulations described in part three of the thesis. It is also worth considering the discussion about the capabilities of the various methodologies outlined in part one and considering whether they can support simulation of the factors described in this section.

As we have already seen, characterization of sprawl is a multi-faceted issue. There are a wide variety of causes that have been attributed to sprawl, however nearly all of the arguments put forward to explain sprawl in North America lack empirical foundation. This is in part due to the diversity of characteristics ascribed to sprawl: people have difficulty understanding exactly what sprawl is, and importantly, in differentiating it from general suburbanization. However, the lack of understanding regarding the causes of sprawl is also a by-product of the fact that there has been relatively work done on the topic, particularly in investigating the *geographic* causes of sprawl. In order to develop simulations of sprawl we need to understand the key factors driving its dynamics, as these will form the rules governing change in a simulation model. This section presents a synthesis of ideas relating to the causes of sprawl, sifting through the literature to pick out some of its key determinants, before using that as a basis for constructing the models described in Chapter 8 and Chapter 9.

In general terms, the sprawling of a city could be perceived simply as one stage in its evolution towards a compact urban structure. Hall (1983) presents a model of urban evolution in which a city passes, at various stages, from a condition of 'primary industrialization' to 'absolute centralization', 'relative centralization', 'relative decentralization', and 'absolute decentralization'. Urban sprawl could be regarded as a function of the latter two stages of evolution in this taxonomy. In fact, the results of simulations in Chapter 8 seem to confirm this idea. Perhaps sprawl is part of a city's natural cycle of development; certainly, the suburbs of the 1950s are beginning to evolve into the sustainable in-town urban communities of today as the city expands. However, simply ascribing the causes of sprawl to urban evolution does little in the way of suggesting mechanisms by which to control it. For that we need to identify more distinct drivers.

Population growth as an engine for sprawl

Demographic transition is one of the most important engines of change in any urban system, and this is also true for sprawl. The expansion of a city beyond its periphery requires, at a minimum, population growth and/or a spatial redistribution of that growth. There are at least three ways in which population growth has contributed to sprawl: absolute growth, relative growth in cities, and restructuring in the dynamics of household demography.

First, cities in North America are—with only a handful of exceptions—growing in terms of absolute population size (Table 2). Even the infamous Detroit metropolitan area, long observed as the poster child for the withering American rust-belt city, has been gaining population on aggregate.

Table 2. A sample of American cities and their population growth (Source: U.S. Census Bureau, Census 2000 Redistricting Data (P.L. 94-171) Summary File and 1990 Census.)

Metropolitan area	**Population (April 1, 1990)**	**Population (April 1, 2000)**	**Percentage change in population**
Las Vegas, NV-AZ MSA	852,737	1,563,282	+83.3
Yuma, AZ MSA	106,865	160,026	+49.7
Naples, FL MSA	152,099	251,377	+65.3
Buffalo-Niagara Falls, NY MSA	1,189,288	1,170,111	-1.6
Bloomington, IN MSA	108,978	120,563	+10.6
Chicago-Gary-Kenosha, IL-IN-WI CMSA	8,239,820	9,157,540	+11.1
Denver-Boulder-Greeley, CO CMSA	1,980,140	2,581,506	+30.4
New York-Northern New Jersey-Long Island, NY-NJ-CT-PA CMSA	19,549,649	21,199,865	+8.4
Seattle-Tacoma-Bremerton, WA CMSA	2,970,328	3,554,760	+19.7
St. Louis, MO-IL MSA	2.492.525	2,603,607	+4.5
Detroit-Ann Arbor-Flint, MI CMSA	5,187,171	5,456,428	+5.2

Second, at the same time, the percentage of the population living in what can be classified as urban areas is also growing (Table 3, Figure 21). Of that urban population, the numbers residing in small cities is swelling at a striking rate. Table 4, which lists the fifty most rapidly growing urban areas (in terms of population) in the United States from 1990 to 2000, features only 13 metropolitan areas with populations exceeding one million, for example.

Table 3. Urbanization over the last hundred years (Source: U.S. Census Bureau.)

Year	Total population	Urban population	Rural population	Percent urban	Percent rural
1900	76,212,168	30,214,832	45,997,336	39.65%	60.35%
1910	92,228,496	42,064,001	50,164,495	45.61%	54.39%
1920	106,021,537	54,253,282	51,768,255	51.17%	48.83%
1930	123,202,624	69,160,599	54,042,025	56.14%	43.86%
1940	132,164,569	74,705,338	57,459,231	56.52%	43.48%
1950	151,325,798	96,486,817	54,478,981	63.76%	36.00%
1960	179,323,175	125,268,750	54,054,425	69.86%	30.14%
1970	203,302,031	149,646,629	53,565,297	73.61%	26.35%
1980	226,542,199	167,050,992	59,494,813	73.74%	26.26%
1990	248,709,873	187,053,487	61,656,386	75.21%	24.79%

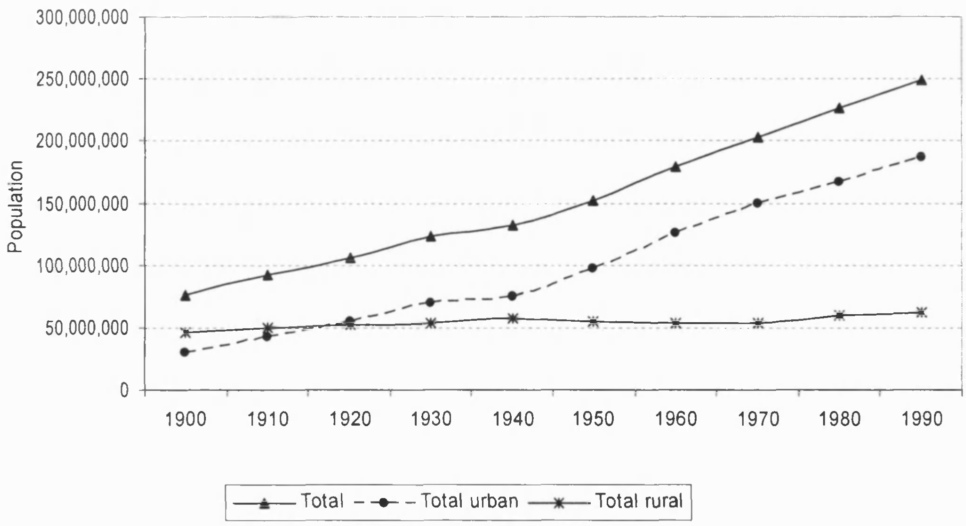


Figure 21. U.S. population growth in urban and rural areas

(Source: U.S. Census Bureau.)

Table 4. Metropolitan areas by percent population change (Source: U.S. Census Bureau.)

Rank	Metropolitan Area Name	Census Population		Change, 1990 to 2000	
		April 1, 2000	April 1, 1990	Number	Percent
1	Las Vegas, NV—AZ MSA	1,563,282	852,737	710,545	83.3%
2	Naples, FL MSA	251,377	152,099	99,278	65.3%
3	Yuma, AZ MSA	160,026	106,895	53,131	49.7%
4	McAllen—Edinburg—Mission, TX MSA	569,463	383,545	185,918	48.5%
5	Austin—San Marcos, TX MSA	1,249,763	846,227	403,536	47.7%
6	Fayetteville—Springdale—Rogers, AR MSA	311,121	210,908	100,213	47.5%
7	Boise City, ID MSA	432,345	295,851	136,494	46.1%
8	Phoenix—Mesa, AZ MSA	3,251,876	2,238,480	1,013,396	45.3%
9	Laredo, TX MSA	193,117	133,239	59,878	44.9%
10	Provo—Orem, UT MSA	368,536	263,590	104,946	39.8%
11	Atlanta, GA MSA	4,112,198	2,959,950	1,152,248	38.9%
12	Raleigh—Durham—Chapel Hill, NC MSA	1,187,941	855,545	332,396	38.9%
13	Myrtle Beach, SC MSA	196,629	144,053	52,576	36.5%
14	Wilmington, NC MSA	233,450	171,269	62,181	36.3%
15	Fort Collins—Loveland, CO MSA	251,494	186,136	65,358	35.1%
16	Orlando, FL MSA	1,644,561	1,224,852	419,709	34.3%
17	Reno, NV MSA	339,486	254,667	84,819	33.3%
18	Ocala, FL MSA	258,916	194,833	64,083	32.9%
19	Auburn—Opelika, AL MSA	115,092	87,146	27,946	32.1%
20	Fort Myers—Cape Coral, FL MSA	440,888	335,113	105,775	31.6%
21	West Palm Beach—Boca Raton, FL MSA	1,131,184	863,518	267,666	31.0%
22	Bellingham, WA MSA	166,814	127,780	39,034	30.5%
23	Denver—Boulder—Greeley, CO CMSA	2,581,506	1,980,140	601,366	30.4%
24	Colorado Springs, CO MSA	516,929	397,014	119,915	30.2%

Rank	Metropolitan Area Name	Census Population		Change, 1990 to 2000	
		April 1, 2000	April 1, 1990	Number	Percent
25	Dallas—Fort Worth, TX CMSA	5,221,801	4,037,282	1,184,519	29.3%
26	Charlotte—Gastonia—Rock Hill, NC—SC MSA	1,499,293	1,162,093	337,200	29.0%
27	Las Cruces, NM MSA	174,682	135,510	39,172	28.9%
28	Brownsville—Harlingen—San Benito, TX MSA	335,227	260,120	75,107	28.9%
29	Richland—Kennewick—Pasco, WA MSA	191,822	150,033	41,789	27.9%
30	Punta Gorda, FL MSA	141,627	110,975	30,652	27.6%
31	Fort Pierce—Port St. Lucie, FL MSA	319,426	251,071	68,355	27.2%
32	Tucson, AZ MSA	843,746	666,880	176,866	26.5%
33	Portland—Salem, OR—WA CMSA	2,265,223	1,793,476	471,747	26.3%
34	Santa Fe, NM MSA	147,635	117,043	30,592	26.1%
35	Houston—Galveston—Brazoria, TX CMSA	4,669,571	3,731,131	938,440	25.2%
36	Bryan—College Station, TX MSA	152,415	121,862	30,553	25.1%
37	Nashville, TN MSA	1,231,311	985,026	246,285	25.0%
38	Grand Junction, CO MSA	116,255	93,145	23,110	24.8%
39	Salt Lake City—Ogden, UT MSA	1,333,914	1,072,227	261,687	24.4%
40	Greenville, NC MSA	133,798	107,924	25,874	24.0%
41	Sioux Falls, SD MSA	172,412	139,236	33,176	23.8%
42	Medford—Ashland, OR MSA	181,269	146,389	34,880	23.8%
43	Daytona Beach, FL MSA	493,175	399,413	93,762	23.5%
44	Springfield, MO MSA	325,721	264,346	61,375	23.2%
45	Killeen—Temple, TX MSA	312,952	255,301	57,651	22.6%
46	Lawrence, KS MSA	99,962	81,798	18,164	22.2%
47	Clarksville—Hopkinsville, TN—KY MSA	207,033	169,439	37,594	22.2%
48	Fresno, CA MSA	922,516	755,580	166,936	22.1%
49	Tallahassee, FL MSA	284,539	233,598	50,941	21.8%
50	Missoula, MT MSA	95,802	78,687	17,115	21.8%

Third, in parallel to a broad trend of population growth, there has been an associated decrease in household sizes and a related increase in the number of housing units (Figure 22, Figure 23).

Of course, if urban populations grow, the city must expand. In some cases, such as that of Hong Kong, this growth has been directed upwards as the capacity of the city to hold population reaches its limits. However, in many cases, even where population densities have grown, the expansion of the city radiates in two dimensions, rather than three, and the extent of the city grows beyond its previous boundary and stretches into agricultural areas or resource lands where possible. In this sense, cities are bound to sprawl by the most general definition of that term—spatial growth—as populations grow. At the same time that urban populations have been growing in absolute terms, however, the distribution of that growth has been concentrated in a spatially distinct manner, largely on the urban fringe. There are a number of possible motivations for this, and in many senses this is where most of the determinants of sprawl lie. We will explore these issues in more detail in Chapter 8, which described a growth-based model of sprawl.

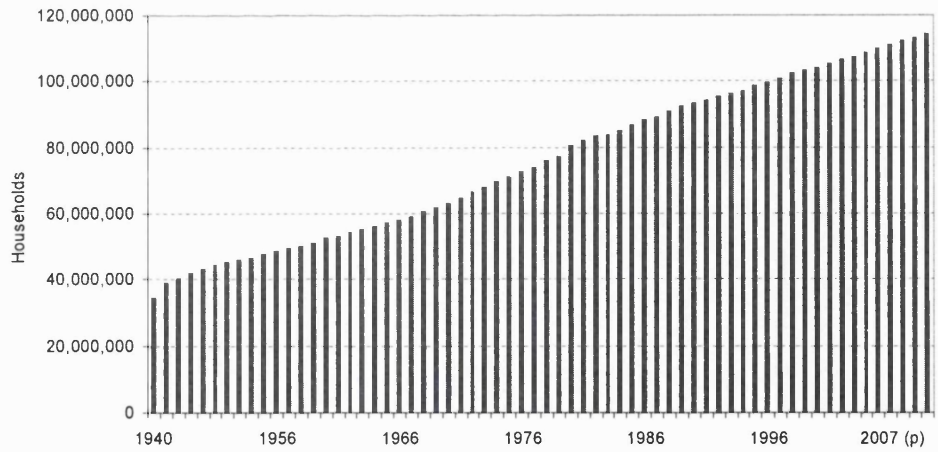


Figure 22. U.S. Household growth since 1940[4]

(Source: U.S. Census Bureau.)

[4] (p) refers to projected figures; figures are projected beyond 2000.

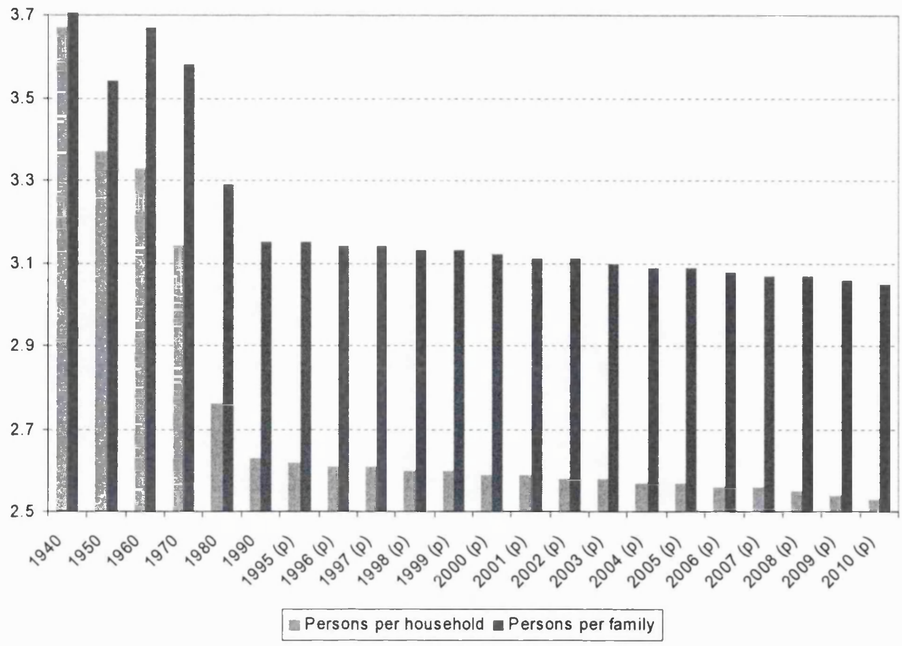


Figure 23. U.S. Household transition[5]

(Source: U.S. Census Bureau 1996.)

Preferences

So, why have urban populations been steadily redistributed toward the periphery? A simple and obvious explanation is that Americans *want* to live in these areas. "[L]ike it or not, the great majority of mankind is praying for [sprawl] to come, to develop and satisfy them" (Gottmann 1967b, p.5). For all the criticisms leveled against suburban living, it is still the preferred living arrangement for an overwhelming number of city dwellers in America (Figure 24). In many cases, people seem to prefer suburban living above other forms. A U.S. survey in the 1990s found that 80% of respondents preferred suburban living above all other types (Morrill 1991). There are a number of likely motivating factors underlying these preferences.

[5] (p) refers to projected figures.

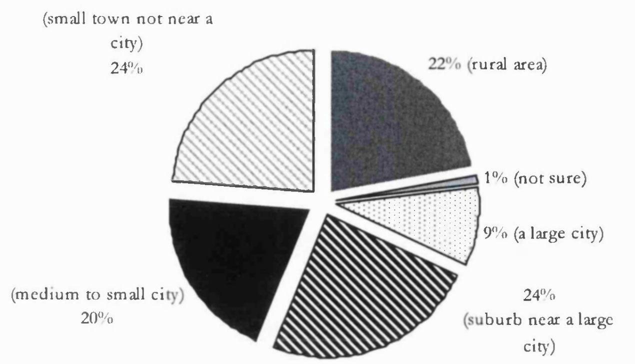


Figure 24. U.S. location preferences

(Source: Fannie Mae Foundation 1997.)

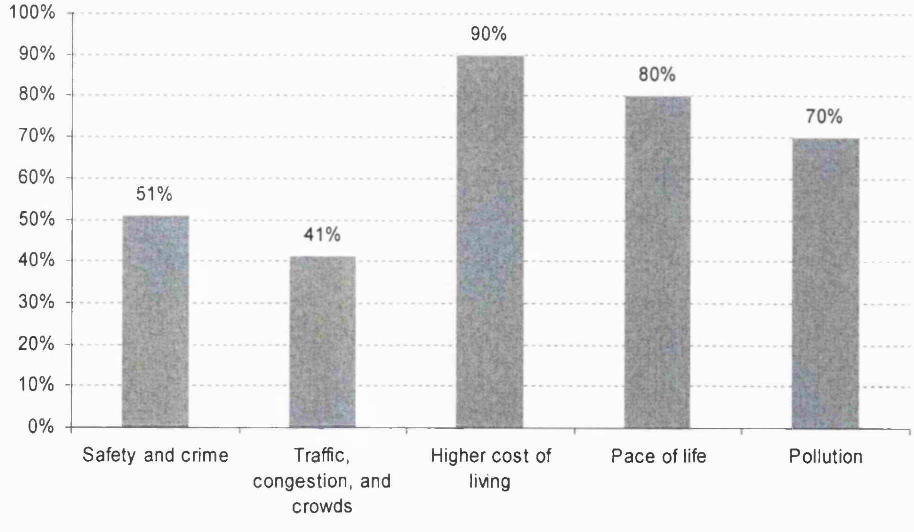


Figure 25. U.S. perceptions of downtown living

(Source: Fannie Mae Foundation 1997; percentages are of those surveyed.)

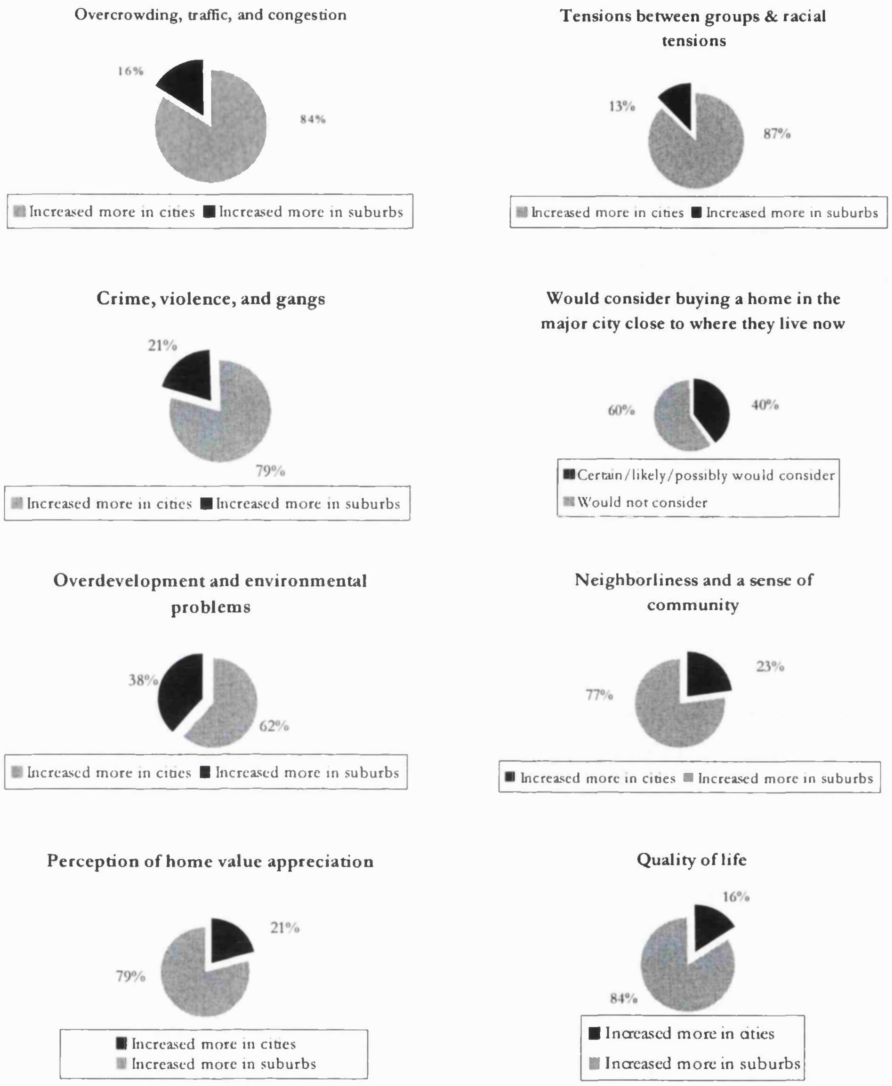


Figure 26. U.S. perceptions of urban living

(Source: Fannie Mae Foundation 1997.)

Some authors have accused outwardly mobile city dwellers of being racially and socially motivated in their decisions to populate the urban periphery with such zeal. It has also been argued that suburban preferences are rooted in a long-standing Anglo-American tradition of ideals based on the exclusion of lower income groups (Audirac *et al.* 1990). In this sense then, much of the motivation for suburban flight could be understood as a negative reaction to the myriad problems of the inner city. To some extent, sprawl has been considered as racially motivated, labeled as 'white flight' (which Pendall (1999)—paraphrasing Farley *et al* (1978)—defines as, "the desire of non-Hispanic whites to move away from minorities, especially African-Americans"). However, there is some debate as to the role of white flight in driving sprawl: studies conducted in the 1980s observed that African-American households were moving to suburban locations in increasing numbers (Farley & Frey 1994). Nevertheless, there is speculation that white households are moving even further out on the urban fringe and into the exurbs (Galster 1991). We will examine the idea that residential preferences contribute to sprawl, from the bottom-up, in the simulations described in Chapter 9.

Public perception certainly paints a bleak picture of city living. There are several apparent disadvantages to life in cities as opposed to the suburbs, including safety and crime, traffic congestion, crowding, higher costs of living, an undesirable pace of life, and pollution (Figure 24). Moreover, public sentiment is of worsening conditions in large cities (Figure 25, Figure 26). In some instances, this is more than perception. The case of Baltimore City is an example. Crime rates in the central city and its neighboring suburban counties have risen since 1985, but at substantially faster rates in the city center (+ 32.6%) compared to the suburbs (+13.4%) (OTA 1995). Aversion to sprawl has not been as vocal as it might have been, largely because sprawl is the physical manifestation of public demands. However, it is important to acknowledge that most people would rather do without the rest of the characteristics of suburban living. Suburbs are in high demand, but are more or less equal in preference with small cities and rural settings (Figure 24).

> "The "American Dream" is not limited to what Frank Lloyd Wright has called the Broad Acres City, just as the British dream is not restricted to the Garden City ideal. There is also the "Mile-high City," to use another of Wright's terms, in the dreams of these nations. And both dreams have long lived together in the minds of most European people. The answer to the planner's dilemma is probably that the citizen's dream is to achieve a mode of life combining all of the advantages of rural setting and rural life, and excluding all of the shortcomings of both" (Gottmann 1967b, p.12).

Decentralization of economic activity

Generally, jobs follow population. For this reason alone, we would expect economic activity to have decentralized from urban cores toward the periphery; businesses generally like to be close to their labor forces. In this sense, population movement has had a pull factor on urban economic activity. However, there are further incentives drawing economic activity to peripheral locations, including lower land and development costs compared to more central locations, and transport networks that facilitate lower costs of movement in outer suburban and exurban locations. There have also been additional push factors driving businesses away from central locations, while the suburbs offer incentives to attract them.

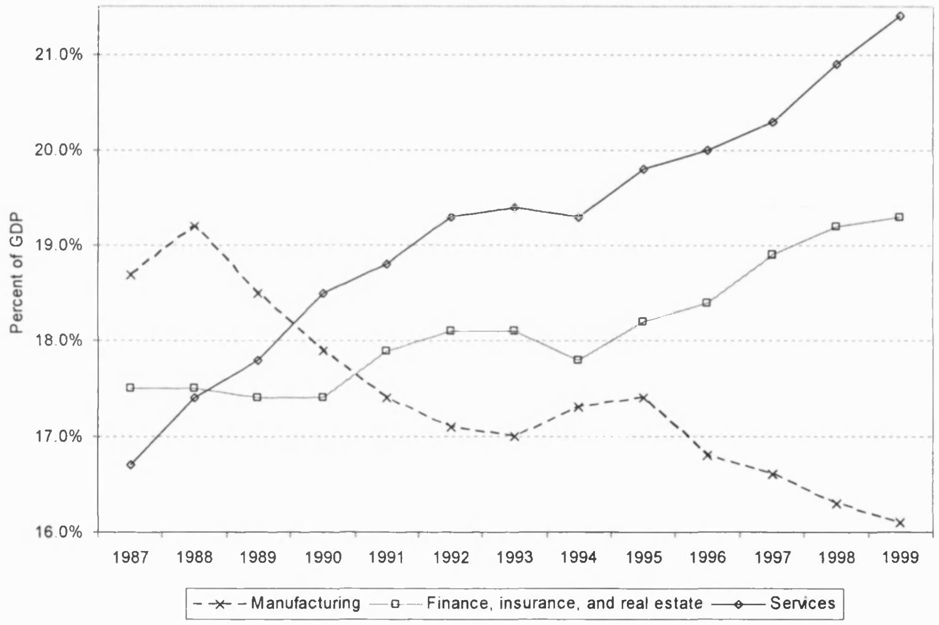


Figure 27. Percent of U.S. G.D.P. by industry

(Source: U.S. Bureau of Economic Analysis, Industry Accounts Data.)

Manufacturing is still productive in many cities, but it now needs fewer workers to maintain a given level of output than it did only a few decades ago (Hall 1983). The U.S. economy is now overwhelmingly dominated by service industries. As Figure 27 demonstrates, manufacturing has steadily declined in its contributions to U.S. G.D.P, while that of the service sector has grown. Accordingly, cities of industry rooted in mechanics and manpower have given way to cities restructured with technology and information. As remarkable as this phenomenon is in itself, the location shifts that have accompanied these industrial restructuring processes have been more interesting.

The Fordist city has been deconstructed and urban areas are now in the grip of "the beginnings of a reconstruction of a new regime of urban development that can be variably called flexible production, flexible accumulation, post fordism, or simply not-Fordism..." (Soja 1995, p.29). These innovations have manifested themselves in the spatial structure of the contemporary city in a number of new ways, notably the rapid growth of sprawling suburbs around declining urban cores. No longer indebted to central cities as interchange points for raw materials and finished goods, industry has diffused rapidly through the suburbs, following its labor forces and pursuing cheap land and easy access to an expanding network of interstate highways in the suburbs. Indeed, most of the new job creation in the city is now in suburban areas (Figure 28). For the most part, only office and office-related jobs remain as dominant sources of employment in central cities and this has not been enough to compensate for losses through the decline of the central manufacturing base. In many cases there is more office space available in non-central locations than there is in the CBD (Figure 29).

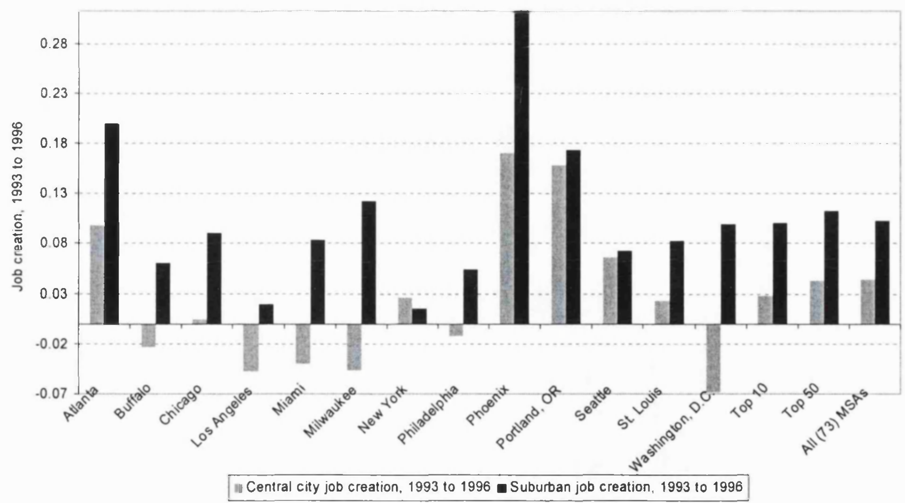


Figure 28. Metropolitan job creation

(Source: HUD 1999.)

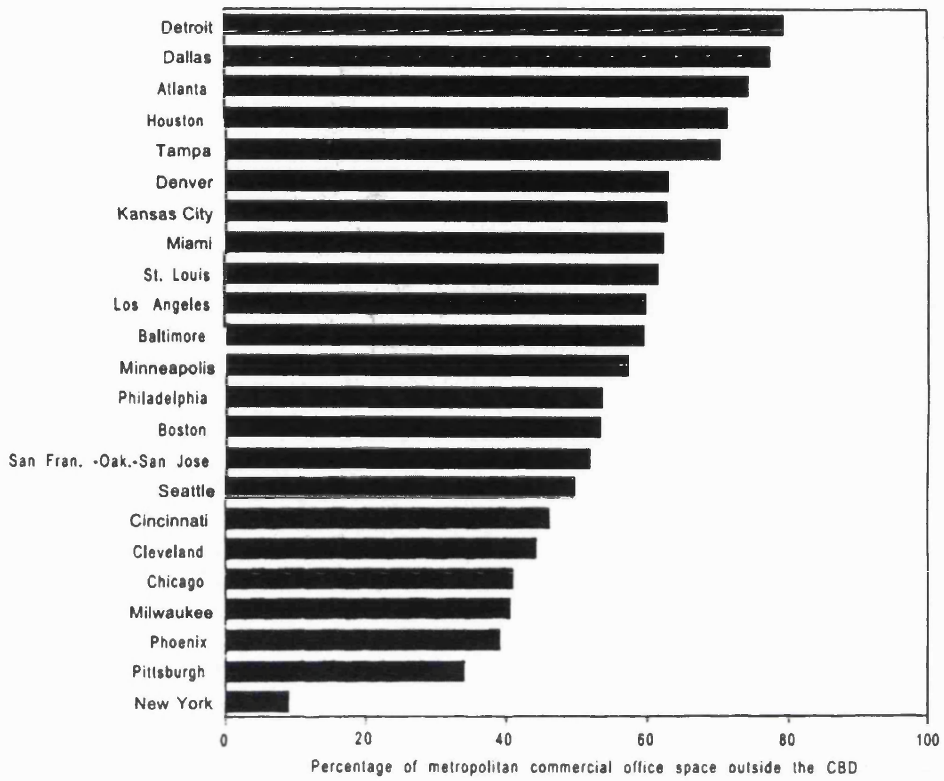


Figure 29. US office space outside the central business district,1988

(Source: OTA 1995, p.197.)

Transportation and telecommunications

After World War II the American interstate highway system was developed with vigor. Indeed, road-building of all varieties proliferated across much of metropolitan America. Accessibility across the city (and between cities) became almost ubiquitous across many urban areas as a result.

> "With development of a cheap ubiquitous transportation system and the decentralization both of residences and of businesses over the past thirty years, accessibility had been greatly increased in US metropolitan areas. The highway system is being well developed, with linkages to the national interstate system as well as to the local system. ... Contemporary metropolitan areas are perhaps better characterized by a homogenous

> activity and transportation surface than by the traditional negative density gradient. ... Given these conditions, then, any number of locations are equally accessible, because locational differences have declined." (Giuliano 1989)

Accordingly, the downtown's pull on location was considerably weakened and gradually highway interchange points became the new center of urban gravity around which development began to orbit. This has manifested itself as "a rapid spread of built-up urban areas with a filling in, at lower densities, of the formerly interstitial sites between the older radial prongs of urban development oriented to rail routes" (Mayer 1967, p.26).

Coupled with changes in the provision of highway infrastructure has been a dramatic growth in the use of the automobile and the dominance of its position in American society. A prolific use of automobiles as a mode of urban transportation makes lower densities possible because it facilitates the dispersion of activities. The operation of an automobile has been greatly facilitated by drops in the price of gasoline in recent decades. The inflation-adjusted price of gasoline in the United States in 1996 was lower than that of 1974 (Gordon & Richardson 1997a). This has encouraged households to substitute housing for transportation costs by moving to suburbs and living farther out, at lower sprawl-type densities.

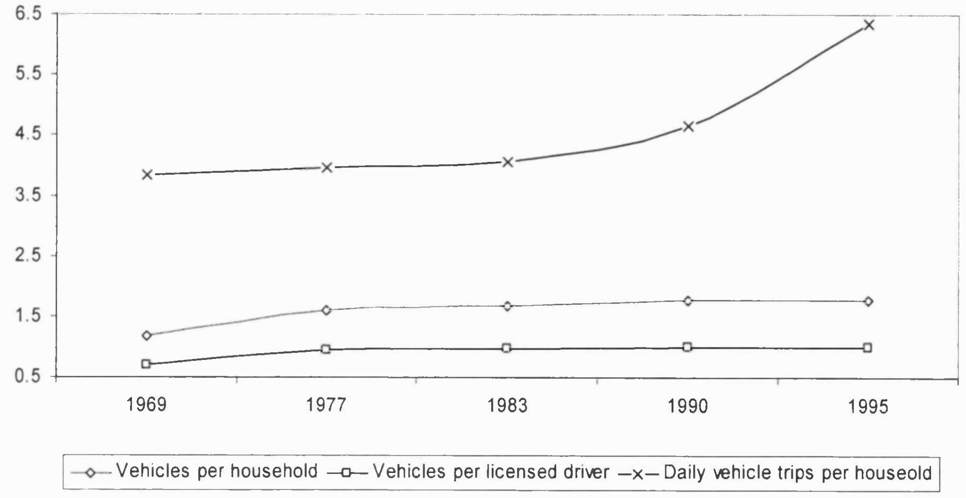
Advances in telecommunications technology have been attributed some responsibility for promoting sprawl. The idea is that modern advances in telecommunications (faxes, email, fiber-optics, cellular phones, etc.) have rendered the downtown clustering of many business activities unnecessary, save those that require face-to-face contact: "The revolution in information processing and telecommunications is accelerating the growth and dispersion of both economic activities and population, possibly moving towards the point where "geography is irrelevant"" (Gordon & Richardson 1997a). Technological advances have greatly extended the "effective radius" of the city—the benefits of agglomeration have been extended over areas of greater spatial extent, allowing many of the costs of congestion to be avoided, and facilitating the rapid sprawl of suburban areas and the diffusion of commercial activity to a scattering of suburban nodes.

A causal relationship between telecommunications technology and sprawl is an easy conclusion to jump to. We do, after all, live in an Information Age, and the transport of information knows little in the way of spatial boundaries: the cover of *The Economist* has already heralded the "Death of Distance" (September 29, 1995). The transport revolution facilitated the growth of the suburbs in the early Twentieth Century by allowing locations to be separated spatially but functionally linked. Information and Communications Technology (ICT) offers the potential to facilitate even further spatial separation between activities.

While a link between the two certainly seems plausible on an intuitive level, evidence has not always supported these arguments. There are three main factors governing the degree to which urban activities can take advantage of telecommunications technology to relocate in urban areas (OTA 1995). The first is the degree to which activities can be transferred electronically into information flows. The second is the degree to which activities rely on spatial proximity to things such as suppliers, customers, competitors, and other units. The third is the degree to which other location advantages remain important. Obviously, this hinges on the activity being considered.

Certainly, telecommunications have influenced sprawl in a positive manner. However, there is also some justification for downplaying the role that ICTs have in driving sprawl. If anything, the clustering of technological infrastructure such as fiber-optic networks in downtowns has had a centralizing influence on urban structure. Furthermore, the notion that city dwellers will substitute telecommunications for transport seems an unlikely one: by most yardsticks of technological penetration, households are now more 'wired' than ever in the United States, yet car ownership levels and VMT continue on an upward trajectory (Table 5, Figure 30).

It is more likely that good old-fashioned vehicular transport is fostering sprawl, rather than the passage of packets and bits through virtual highways. Specifically, developments in the interstate highway network and changes in the use and proliferation of automobiles in urban areas have served as important catalysts for sprawl.



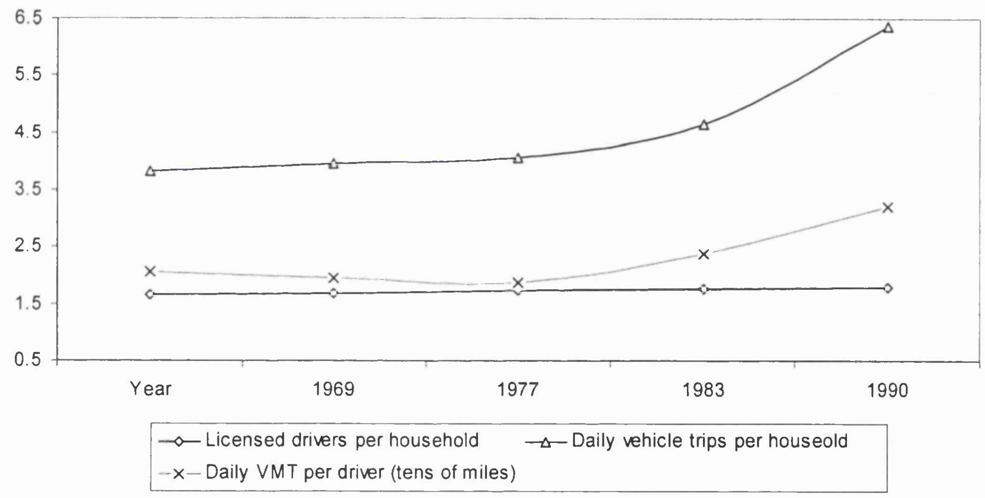


Figure 30. Changes in U.S. travel behavior

(Source: FHWA 1995.)

Table 5. U.S. travel behavior (Source: FHWA 1995.)

Year	1969	1977	1983	1990	1995
Licensed drivers (mil.)	103	128	147	163	176
Household vehicles (mil.)	73	120	144	165	176
Household vehicle trips (mil.)	87,284	108,826	126,874	158,927	229,745
Household VMT (mil.)	77,594	907,603	1,003,139	1,409,600	2,068,368
Person trips (mil.)	145,146	211,778	224,385	249,562	378,930
Person miles of travel (mil.)	1,404,137	1,879,215	1,946,662	2,315,300	3,411,122
Vehicles per household	1.16	1.59	1.68	1.77	1.78
Licensed drivers per household	1.65	1.69	1.72	1.75	1.78
Vehicles per licensed driver	0.7	0.94	0.98	1.01	1
Daily vehicle trips per household	3.83	3.95	4.07	4.66	6.36
Daily VMT per driver	20.64	19.49	18.68	23.69	32.14
Average vehicle trip length (miles)	8.85	8.34	7.9	8.98	9.06

Developers

Developers have been blamed for encouraging scattered development in expanding suburban districts of North America. For the most part in growing cities, real estate competitors act independently in their development decisions. This promotes a discontinuity in the spatial pattern of their developments. More specifically, the independence of development decisions and their lack of coordination in suburban construction encourage speculation—the withholding of land for development. Because of speculation, large areas of land in the suburbs may become priced out of any market save urban usage, but yet may still be withheld from use (Clawson 1962). There are a number of factors motivating the decision of landowners to withhold their land from development. Preferences differ among landowners; some may need to sell their land quickly, while others may elect to hang on to it in the anticipation of gaining a higher price for it at a future date. In addition, there are many institutional elements that might foster speculation, including estate holdings, trusts, defective titles, and covenants (Clawson 1962). Also, Pendall (1999) argues that fragmentation in the ownership of agricultural land results in the piecemeal release of sites for development. Each of these factors contributes to the leapfrogging of fringe urbanization and a sprawl-like pattern of development.

The practices of financial lending agencies are also responsible for promoting the fragmented nature of development on the periphery. To minimize risks and to shield themselves from losses, lending groups may avoid committing to lone builders in single geographic areas. In doing so, creditors insulate themselves from a concentration of risks in a single project, builder, or area (Harvey & Clark 1965). For similar reasons, lending institutions may favor projects that have a specific completion timeframe within a limited number of building seasons (again because of factors of risk and return). Once again, this type of lending practice favors discontinuous, incremental-style sprawl development and fosters scattering.

Planning

American suburban sprawl is in many senses, very much a mindset inherited from earlier ideas about the city that planners espoused. The suburban ideal, and city dwellers' preferences for it, is part of an intellectual tradition in planning history,

> "instigated by the Garden City movement, from Ebenezer Howard to Clarence Stein and Lewis Mumford, with their "devotion to the idea of a cozy town life," and instigated as well by Le Corbusier's utopian city of "towers in the park"...urban visions [that] had merged into a particular doctrine of city planning that on the one hand attacked high-density, concentrated urban forms, replacing them with low-density suburbs, and on the other hand, selectively dismantled the intricate fabric of central city life with zoning and urban renewal programs" (Audirac *et al.* 1990, pp.472-473).

American land-use planning policies may have fostered sprawl, at least indirectly. When applied spatially with varying degrees of enforcement, land-use controls can create an imbalance in the attractiveness of competing areas. If there is a discrepancy between controls inside and outside of a city's boundary, for example, land-use planning often makes the lesser controlled area more attractive, i.e., the urban fringe

(Harvey & Clark 1965). Often a regulatory body may not have control over an entire housing market area. If this causes building and land-use standards within the controlled area to be more stringent than common practices in development, those standards themselves may induce construction outside of the controlled area, thus contributing to sprawl (Harvey & Clark 1965). In a comprehensive empirical study in the United States, Pendall (1999) found that low-density-only planning policies and building-permit caps actually foster sprawl, as well as exacerbating racial and ethnic divisions.

Barnett (1995) argues that outdated planning regulations are responsible in large part for sprawl. Commercial strips, a 1920s planning idea, were a form of design that was built into thousands of American suburban ordinances that were put into place in the 1950s, however, "apparently no one stopped to contemplate the effect of mapping commercial land exclusively in narrow strips along highways where the only means of access was the automobile" (p. 47). Outdated planning regulations are also responsible for much of the formless nature of sprawl and its sameness of design. Suburban-style regulations were designed to fit relatively small increments of new development into already established urban areas. Lot-by-lot zoning and subdivision was not intended to become the only development control over expansive geographic areas (Barnett 1995). Some authors also blame urban planning containment policies for perpetuating sprawl (Audirac *et al.* 1990). They argue that development restrictions (such as urban growth boundaries) that make individual communities more compact actually have the counter-effect of shifting growth to less restrictive exurban sites.

Municipalities, for the most part, have failed to confront residents in the suburban periphery with the full costs of providing public services (schools, disposal, utilities) in sprawled areas of the city (Ottensmann 1977). In this sense, urban service provision acts as a subsidy for sprawl because services are priced independently of their distance from central facilities (Ewing 1997). Federal funding of waste treatment facilities in the 1960s, 1970s, and early 1980s, for example, encouraged suburban sprawl by providing the services that enable city expansion to take place. Also, the Environmental Protection Agency (EPA) has permitted and encouraged the construction of large and extensive sewerage systems in sprawling areas of the city (Ewing 1994).

Public policy

> "[I]t is widely believed that a myriad of government policies, including tax policies, depreciation allowances, building regulations and implicit subsidies, subsidize sprawled Greenfield development and discourage efforts to reuse older urban and suburban land and infrastructure. ... There are also sets of regulations that potentially contribute to sprawl. These include the Americans with Disabilities Act and laws aimed at health and safety at work, which make it less costly to build an entirely new building than to buy an existing building and bring it up to the standards demanded by these laws." (OTA 1995, p.196).

Bahl (1968) generalizes the role of tax policies in fostering land sale speculation as being the function of a number of factors: the optimum time period for withholding land from development, an individual landowner's interest rate, the discounted expected percentage return on an alternative investment, the net rate of return on the land, the discounted expected return on the land, the marginal personal income tax rate of the land owner, the number of years in the landowner's speculative time horizon, a subjective factor reflecting risk, and property taxes. Considering these factors, then, it is clear that a number of important aspects of tax policy weigh heavily into a landowner's decision whether or not to withhold land from development, causing sprawl.

American tax policies have been equally accommodating of urban sprawl, by favoring home-ownership over renting. The costs of home-ownership are subsidized by U.S tax policy in the form of several income tax deductions: mortgage loan interest, capital gains tax deferment, and property tax payments (OTA 1995). It could also be argued that income tax deductions for mortgage interest payments generally favor suburban residents over all other taxpayers in the city—there is an indirect geographical bias to Federal tax policies. Tax policies favor new homes and single-family housing: exactly the type of development that is overwhelmingly represented by sprawling areas of the city.

> "It is generally agreed that in the past the public sector encouraged low-density suburbanization through tax deductions, mortgage guarantees, and depreciation formulas favoring new construction over the upgrading and repair of existing structures. That is, dispersed urban development was encouraged by large implicit subsidies for homeownership and single-family housing because, as Peterson notes: "The new, low-density construction favored by tax laws is obviously most suitable for location outside the central metropolitan core." Though the spatial implications of the federal tax code have not been studied more recently, it is reasonable to conclude that this subsidy continues to sponsor sprawl." ((OTA 1995; Peterson 1980, p.200)

The federal tax code also emphasizes the creation of subdivisions in small and discontinuous increments. Land is commonly sold to developers in installments so as to minimize capital gains on their income tax returns. In addition, sub-dividers and developers may limit their programs for any taxable year so as not to slip into higher tax brackets which might incur increased rates of taxation on their profits (Harvey & Clark 1965).

In some cases, public policy has had a more direct influence in encouraging suburbanization. Many state incentives (free land offers, subsidized training, tax breaks, tax exempt industrial development bonds, low interest loans) are biased against central cities. There have been well-documented instances where policy has been used to intervene at the state level to facilitate the suburbanization of large employers: $110 million in subsidies in Illinois for the relocation of Sears to suburban Hoffman Estates, for example (OTA 1995).

Political fragmentation

Some authors contend that political fragmentation in American governance has exacerbated sprawl. The Chicagoland metropolitan area stretches over 3800 square miles and includes 265 municipalities, 1200 individual tax districts, all distributed over six counties in three states (OTA 1995). The argument here is that the plurality of jurisdictional units in American cities, each with varying levels of autonomy, makes it difficult to coordinate urban growth in a sustainable manner. Also, the competition for property and sales tax revenues between jurisdictions may aggravate sprawl in some cases (Orfield 1997). However, it seems intuitive that fragmentation might permit diverse preferences for living within a metropolitan area, thereby reducing the likelihood of sprawl. Obviously, this is less likely to be the case if there is not diversity along socioeconomic lines.

Conclusions

This chapter has reviewed the literature pertaining to sprawl, synthesizing a diverse range of opinions regarding the proposed characteristics, costs, and causes of American sprawl and offering empirical data to evaluate the validity of several of these claims. Sprawl is a very significant phenomenon, perhaps representative of a new phase of contemporary urbanization, fueled by contemporary preferences for urban living that go back to the days of Ebenezer Howard, but are now manifest in new ways because of contemporary transport infrastructure. The literature on sprawl's characteristics and causes is still at a relatively exploratory stage. Simulation tools have much potential for evaluating and testing the ideas discussed in this chapter, and this is done in part three of the thesis.

A number of characteristics have been suggested for identifying sprawl "on the ground". These include density, scattered attributes, open space, aesthetics, and accessibility. In addition, it is important to recognize that sprawl is a very dynamic phenomenon and these characteristics may change rapidly with time.

The consequences of sprawl have been investigated by a number of authors, both from pro-sprawl and anti-sprawl perspectives (but mostly from an anti-sprawl standpoint!). Sprawl does, actually, have a number of benefits, although these are popularly overlooked. 'Costs' of sprawl can be grouped into direct costs, mostly related to infrastructure, and indirect costs. Indirect costs include air pollution, downtown decline and associated spatial mismatches between central cities and suburbs, loss of agriculture and natural resource land, and erosion of water quality and general quality of life.

The causes of sprawl are relatively difficult to isolate, simply because *so much* contributes to the phenomenon. Nevertheless, we can identify a few key drivers, including population growth and demographic transition, preferences, decentralization trends in economic activity, transport and telecommunications, developers, planning regimes, public policy, and political fragmentation.

Chapter 7 discusses how these ideas might be conceptualized in a model designed to explore the geographic dimensions of sprawl. Chapter 7 also explores the literature on modeling sprawl, although there have been relatively few studies targeting sprawl specifically. The discussion in this chapter also serves as the foundation for the design of simulation experiments in part three, using methodologies based on the automata, cellular automata, multi-agents systems, and Geographic Automata Systems outlined in Chapter 4 and Chapter 5, as well as incorporating elements of the methodologies mentioned in Chapter 2. Several of the characteristics of sprawl mentioned in this chapter feature in those models, either as variables or outputs. The discussion of causes of sprawl also serves as the basis for various rules, engines, and entities in those later models.

Chapter 7. Toward a model of suburban sprawl

"And looking backward is very nearly as much fun as looking forward, though our digital soup does thin out rather rapidly, that way down the time-line." (Gibson 1999, p. 251)

Introduction

The previous chapter discussed suburban sprawl—its characteristics, consequences, and potential causes. This chapter reviews the literature on urban models in a sprawl-specific context, assessing a number of simulation environments in their capacity to serve as sprawl models and analyzing their features to determine ways in which the attributes discussed in the previous chapter might be represented in a modeling context.

Despite the importance of sprawl in the public and political agendas in the United States, a comprehensive sprawl model has yet to be developed. Many models touch on sprawl-related issues or feature *elements* that might also be used to explore sprawl phenomena, but for the most part these models have been built for other purposes. The term "sprawl" features in much of that literature, but only in a very general context, as a synonym for general urbanization or rapid growth. The research landscape for sprawl simulation looks something like this: a handful of models developed to deal with sprawl-like mechanisms of urbanization (polycentric cluster formation and rural-to-urban land conversion), and a body of literature replete with references to general urbanization.

That is not a criticism; sprawl is still a relatively poorly understood phenomenon and the models that will be discussed in this chapter were designed to serve uses other than sprawl simulation, and serve those alternative uses well. Nevertheless, we *can* identify several of the features discussed in the last chapter and it is useful to explore their representation in previous work before introducing the simulations in part three of the thesis.

The next section re-visits the discussion in the last chapter, this time in the context of methodological requirements for a sprawl model: general requisites for dynamics, geography, and inter-dependence among sub-systems; likely model variables; and necessary processes. An analysis of pertinent urban simulation literature follows that discussion, focusing on density models, general planning support systems, general urban growth automata, models of polycentricity, models of fringe urbanization, and their ability to satisfy the aforementioned requirements.

Requirements for a sprawl model

Chapter 6 outlined the broad range of characteristics, consequences, and causes associated with American suburban sprawl. From this 'soup' of attributes, we can identify some of the key components necessary for constructing a model of sprawl. Of course, not everything can be simulated! Several of the properties mentioned in the last chapter do not lend themselves to empirical measurement, e.g., aesthetics and some of the indirect costs connected to sprawl. It is difficult to model these components, because of their intractability. Nonetheless, we can, perhaps, identify some of the core components required for building robust sprawl models.

General requirements

As mentioned in Chapter 6, one of the key characteristics of suburban sprawl is its dynamic attributes. American sprawl is a rapid phenomenon, voracious in its appetite for land. It is also an evolutionary phenomenon. Sprawled areas of a city may develop into quite sustainable urban areas with time, as large single-lot land parcels become sub-divided and developed at higher densities, and previously fragmented areas are subjected to infill. Nevertheless, sprawl is commonly treated as a static phenomenon in analysis, its properties captured or modeled in snapshots. A sprawl model should include dynamics as a fundamental component; it should facilitate the representation of sprawl as a dynamic, *evolving* process. This necessitates the use of simulation methodology that permits representation of evolutionary dynamics. As discussed in Chapter 2, much of the traditional modeling methodology in common use for spatial simulation is incapable of supporting evolutionary dynamics. The automata tools outlined in Chapter 4, however, are ideal for building models of evolutionary systems. Automata tools, with their emphasis on interaction, are also excellent for treating sprawl as a complex adaptive system and for evaluating emergence in terms of evolutionary dynamics.

There are also spatial requirements for sprawl models. Many of the characteristics, consequences, and causes of suburban sprawl are non-spatial in nature. The literature on sprawl reflects this; much of the published work on sprawl is in fields outside of geography, and much of the literature overlooks the role that geography plays in the phenomenon. Nevertheless, geography does matter! The end-result of all sprawl mechanisms, whether economic, social, political, etc., is that they manifest in particular *places*, with particular *spatial configurations*. Sprawl occurs in some locations but not others. Some cities and urban areas are compact and 'sustainable'; others sprawl. Space is therefore a fundamental requirement for sprawl models. Once again, this has methodological implications. Some of the conventional modeling methodologies are relatively weak in their treatment of space. In particular, they can be regarded as comparatively ineffectual in terms of connecting spatial patterns to the geographic (or non-geographic) processes that generate them. They are also relatively vapid in their capacity to handle phenomena that operate across multiple scales. Automata models, by comparison, are excellent tools for connecting form and function, pattern and process. They can also handle multi-scale phenomena relatively seamlessly.

Interdependence of sub-systems is yet another important requirement for sprawl models. The literature review and discussion about sprawl in Chapter 6 is broad, drawing from several fields of research. This is because sprawl is an all-inclusive phenomenon, with characteristics, consequences, and causes that span multiple domains: physical, social, economic, political, ecological, environmental. Representing the complex inter-dependencies between these diverse systems and sub-systems is a complicated endeavor. Even simulating the *geographic* influence and domain underpinning the systems is an intricate task. Nevertheless, these inter-dependencies are an important component of sprawl and should feature in a sprawl model. Many of the methodologies discussed in Chapter 2 are less than ideal for supporting a simulation with several co-dependent sub-systems. Automata tools, on the other hand, offer a pliancy that renders them more suitable to multi-process modeling, and perhaps more appropriate to simulating sprawl. An almost limitless range of variables may feature in an automata model, and almost any process may be represented by transition rules.

Sprawl characteristics and model variables

The selection of appropriate variables for the sprawl models in part three of this thesis is an important consideration. A wealth of characteristics may be attributed to sprawl. Many of these could, potentially, feature in a model of sprawl. However, one of the requirements for characteristics introduced as variables in a sprawl model is that they be empirical. Also, for the sake of parsimony (and tractability) it is wise to use only the *key variables* necessary for explaining the phenomenon.

Much of the discussion surrounding sprawl characteristics is descriptive in nature. Yet, empirical measures for identifying and measuring sprawl can be identified (Torrens & Alberti 2000). Selecting the key variables for a sprawl model is somewhat more contentious. Several variables play a role in the formation of sprawl, and there is little understanding about their relative importance. Indeed, one of the purposes of a sprawl simulation might be to determine the relative contribution of various characteristics to the phenomenon. It is perhaps appropriate, then, to include multiple variables representing sprawl characteristics, and to look at their interaction in a simulation. Following from the discussion of sprawl characteristics and consequences, we can identify several key empirical variables that should feature in a comprehensive model of sprawl.

The *minimal* set of variables would likely include density (of population, employment, or some other activity variable); spatial distribution of urban extent; patterns of urbanization; some determination of the functionality of urban space; aesthetic characteristics (although these are very often subjective in nature and very difficult to quantify); and accessibility. Several of these variables will be important in generating model outcomes. Many are spatial in nature and others necessitate visual representation in a simulation. Taken together, the set of variables listed above would seem to discount many of the conventional simulation methodologies discussed in Chapter 2. The flexibility of the state variable concept in automata modeling, and the ability to represent multiple states in a related lattice or network makes automata tools an ideal mechanism for simulating sprawl characteristics.

However, it is not feasible to model all of these variables; the models discussed in part three therefore focus on a subset of the characteristics discussed in the previous chapter.

Sprawl drivers and modeled processes

A number of potential devices could be introduced as drivers in a sprawl model. Because the factors responsible for sprawl are not fully understood, any mechanisms that feature in a sprawl model will be, to a certain extent, exploratory in nature. The goal of a sprawl model could in fact be to test a spectrum of potential causes and their relative importance in explaining sprawl characteristics. This is what we will do in part three; again focusing on a subset of the causes discussed in the previous chapter.

At a very broad level, a sprawl model should be capable of representing some key processes. First, overall system growth is important. It sets the general pace for the model. Second, mechanisms should be introduced to explain the distribution of that growth over space. There are a number of potential processes responsible for explaining the geography of sprawl. Adequate representation of sprawl drivers will require representation of processes operating on a very general level, at a city-wide scale. It also necessitates the introduction of components designed to mimic the actions of entity-level units of the urban system, such as developers and individual households. Third, there should be some representation of institutional influences, such as public policy and planning.

Automata tools are flexible enough to be suitable for representing just about any process in a sprawl model. However, there may be some processes that operate from the top-down that are not suitable to automata-based simulation.

The research landscape for urban sprawl modeling

As mentioned in the introduction to this chapter, there is relatively little literature on the subject of sprawl simulation. However, several models deal with sprawl-like mechanisms or model elements of sprawl phenomena. This section reviews that literature base, examining previous work in the context of the requirements outlined in the previous section.

Density models

Density is one of the key components of sprawl. A vast literature devoted to urban density modeling exists. In particular, there has been a lot of research into urban density gradients. Authors have looked at the decay in density of a variety of urban activities with distance from a central core or multiple sub-centers (Alperovich & Deutsch 1992; Bussière 1968; Mills 1972; Muth 1969; Zielinski 1979). Considerable attention has been afforded to the formulation of decay functions for describing density decline: inverse power functions (Smeed 1963), negative exponential functions (Batty & Kwang 1992; Bleicher 1892; Clark 1951), and equilibrium functions (Amson 1972, 1973). Despite some work loosely relating density gradients to explanatory *variables* accounting for factor substitution between consumer preferences and house prices (Mills & Tan 1980), and edge city formation (Benguigi *et al.* 2001; Wang & Zhou 1999), this work has focused on *describing* density gradients rather than *explaining* the mechanisms that generate them. The actual processes that give rise to particular densities are relatively weakly explored.

Planning Support Systems

Some popular and innovative planning support systems (PSS) have been developed in very recent years. To a certain degree, these systems go some of the way toward bridging the gap between conventional modeling approaches (such as those mentioned in the last section and discussed in detail in Chapter 2) and automata-style methodologies. Three PSS stand out, in particular—the California Urban Futures family of models, the What if? system, and UrbanSim. None of these PSS are designed to simulate sprawl, although UrbanSim comes particularly close. However, they do touch upon several of the key requirements mentioned earlier in this chapter and are noteworthy in being *almost* automata-like, with associated benefits in relation to model dynamics, geography, and inter-dependency.

The California Urban Futures series of models are being developed by John Landis and colleagues at University of California, Berkeley's Department of City and Regional Planning. The series consists of the California Urban Futures (CUF) Model; its successor, California Urban Futures Model II (CUF II); and the California Urban and Biodiversity Analysis Model (CURBA) (Landis 2001). All three were designed for use in practice as PSS.

The CUF family is relevant to this discussion in a few ways. First, CUF II is automata-like in so much as it deals with grid-cell-based partitions of urban space and incorporates transition-like rules in the form of logit models. Second, the models are designed for scenario exploration, and some of the supported scenarios touch upon sprawl-related issues.

CUF consists, essentially, of a series of economic models tied to GIS (Landis 1994, 1995). CUF II is more decision-based, treating land-use transition as a series of nested logit models (Landis & Zhang 1998a, b). CURBA is designed to explore environmental issues through habitat analysis (Landis 2001). The CUF and CUF II models work in a modular fashion. CUF contains components for general population projection, a model for assigning growth to locations, and decision rules for land annexation. CUF II works in a similar way, but assigns growth probabilistically, based on a series of logit equations that incorporate surrounding land-use, among other factors, as an explanatory variable.

Ultimately, the CUF series are land price models—phenomena are modeled in relation to their utility in a simulated land market. However, the models do feature several of the factors we consider important in explaining sprawl: growth, activity (accessibility), and public policy (CUF II features a proxy variable for policy: an assessment of whether or not a site is within a policy sphere-of-influence).

The What if? Planning Support System is being developed by Richard Klosterman and colleagues at University of Akron, Department of Geography and Planning, and Community Analysis and Planning, Inc. It is designed as a PSS with a policy emphasis. The model functions on the premise of simulating what might happen if certain policies were introduced, or certain assumptions were upheld in the evolution of a city.

The What if? System is relevant to the sprawl discussion because of it strong policy rationale.

> "*What if?* is most appropriate for areas that are experiencing, or anticipating, rapid urbanization and the associated problems of traffic congestion, inadequate public infrastructure, and the loss of agricultural and open land" (Klosterman 2001) (p.267).

Clearly, these properties mirror much of the discussion about the consequences of sprawl in Chapter 6.

What if? works in much the same way as the CUF series. It is modular in design, with separate components handling population growth, land supply, and land demand (Klosterman 1999). Growth and change are allocated to geographies above parcel level and these can be aggregated to produce regional-level attributes. A supply ('suitability') model determines available land supply in a simulation and its ability to accommodate future demand. A 'growth' model determines demand for that land. An allocation model then assigns demand to suitable locations. Supply and demand are calculated based on a set of user-defined weights and constraints.

The model features many of the variables we consider necessary for simulating sprawl, including density, accessibility, and open space. In addition, it supports a range of sprawl-relevant policy scenarios, including agricultural land and open space protection programs. It also features options for specifying preferred growth patterns, although this refers to general urban structure (radial, concentric) rather than sprawl-specific pattern types.

UrbanSim is being developed by Paul Waddell and colleagues at the University of Washington's School of Public Policy, Department of Urban Design and Planning, and Department of Computer Science. UrbanSim is designed for use as a PSS, and has been applied in practice in Oregon, Hawaii, Washington, and Utah. The model is of relevance to this discussion for a number of reasons. It is strongly behaviorally-rooted. Model algorithms emphasize preferences in determining land and real estate market dynamics. It is therefore agent-based in some senses. The level of geography is very fine-scaled; individual land parcels can be represented. UrbanSim is also a micro-simulation model, disaggregate in its treatment of urban actors. It is also longitudinal in its treatment of time, closely approximating dynamics in the sense of the automata models discussed in Chapter 4.

UrbanSim works around a set of integrated models (Waddell 2000, 2001, 2002; Waddell *et al.* 2003).

- A macroeconomic model handles regional forecasts externally and sets up control totals for model variables.
- A land price model determines the price of land at each land unit in a simulation. This model is formulated using hedonic price equations that include neighborhood-like attributes.
- A development model is responsible for simulating development choices and decisions about what to develop and where to build. The development model establishes a list of development alternatives for each year in a simulation and assigns them a probability of being developed, using a multinomial logit model.
- An accessibility model predicts patterns of accessibility by auto ownership.
- A demographic transition model is responsible for forecasting births and deaths in the simulated population.
- An economic transition model simulates job creation and loss.
- A household mobility model determines whether households decide to move or not in a simulation. The probability of movement is based on historical data.
- An employment mobility model simulates which employers (or 'jobs') will move.
- A household location choice model is responsible for simulating relocating households' choice of location, based on preferences for housing.
- An employment location choice model simulates the choice of location for 'jobs', again based on preferences.

UrbanSim also links to travel models and environmental models (Alberti & Waddell 2000), as well as connecting to a variety of scenario-based assumptions. Available scenarios include the imposition of urban growth boundaries for containment, support of transport-induced development, general polynucleation and polynucleation connected by multi-modal transport infrastructure, and the encouragement of development in relatively impoverished areas (Waddell 2001).

UrbanSim has all the ingredients necessary for exploring sprawl scenarios. Most of the sprawl characteristics discussed in Chapter 6 feature in the model. However, what distinguishes it from the other PSS discussed in this section is the emphasis that is placed on *decision-making*. UrbanSim thus handles many of the *causes* of sprawl, explicitly, as urban processes in their own right, albeit with an over-arching emphasis on market forces rather than geography.

General urban growth automata

Automata-based models of general urban growth are a particularly popular thread in urban simulation research. These models offer several advantages over more conventional approaches such as those outlined in the last section. (Although the boundary between conventional and automata-type approaches is dissolving, as the previous discussion about PSS highlights.) In particular, the emphasis on bottom-up, dynamic, local-scale, interactive formation of urban phenomena is particularly relevant in the context of sprawl. These qualities help to satisfy several of the general requirements of a robust sprawl model. However, very few automata models have been developed for sprawl applications.

The Dynamic Urban Evolution Model (DUEM)

Michael Batty and Yichun Xie have built a series of CA models of general urbanization using their Dynamic Urban Evolution Model (DUEM) package. Their model is based on the premise of cities going through various lifecycle stages: birth, growth, and decline; or 'initiating', 'mature', and 'declining' (Batty *et al.* 1999b). The model determines the proportion of activity (development) in each of these stages at each transition point in a simulation. A variety of rules govern the proportion of activity in each stage, through determination of development potential for cells. Potential is contingent upon a number of factors, including distance from existing development (with distance-decay specified using one of a range of decay functions), direction to existing development (to accommodate sector- or wedge-like growth), and the intensity of development (number of developed cells) in a neighborhood. In addition, various land-uses are introduced as state variables, with user-defined constraints on the probability of transition between them. Transport infrastructure also features, with roads grown by diffusion-limited aggregation (DLA) (Batty *et al.* 1989). Other applications were developed to explore the roles of feedback (time-lag) and innovation ('noise') in urbanization (Batty 1998).

The model features several components relevant to sprawl. The classification of urbanization in life cycles is interesting. However, the final stage is characterized as decline, whereas we might determine sprawl as a potential candidate instead. The use of distance-decay has particular significance in terms of accessibility. Also, the simulation of road *development*, rather than simple use of roads as an influencing variable, is particularly innovative. It introduces the idea of road-building as an integral part of the urbanization process. Of course, the development of transport infrastructure on the fringe of cities, in particular, is very instrumental in opening up peripheral areas to sprawl-like growth, as was mentioned in Chapter 6.

Wu and Webster's models

Fulong Wu and Chris Webster built a model of urban growth, focused on growth through development. The model was applied to the Guangzhou region of China and was designed to test how decision-making affects urban growth (Wu & Webster 1998). Other applications saw simulations developed to explore property rights and regulation regimes (Webster & Wu 1998; Wu & Webster 2000).

In the Gunagzhou example, transition rules were derived from a multi-criteria evaluation technique that allows responses from decision-makers (or game theory models of decision-making) to be used in the formation of rules. Development was based around the determination of probabilities for land conversion, from resource land to built-up uses. The derivation of that probability is determined using a number of factors. Several of these relate to sprawl. Accessibility is particularly well-represented, with factors including cost of travel to a city center and a major industrial district, access to a railway station, and access to highways. In addition, topographic and regulative restrictions feature in transition probability calculations.

The Queensland models

Ward and colleagues developed a series of CA models of urbanization in Queensland, Australia, applying them to the simulation of urban growth in Brisbane and the Gold Coast area (Ward, Murray *et al.* 2000) as part of a project monitoring urban growth in rapidly urbanizing areas (Ward, Phinn *et al.* 2000). They were particularly interested in introducing constraint parameters to control simulation runs, and several of those parameters resemble the sprawl characteristics discussed in Chapter 6.

Simulations begin with population projections and growth is assigned, probabilistically, to cells, based on an array of dependent variables. The authors paid particular attention to accessibility variables. Their model of Brisbane explored accessibility and its role in determining urbanization in particular detail. Their results demonstrated the importance of overall system connectivity in generating realistic growth models. They found that it was not sufficient to specify accessibility only in local, cell-by-cell, terms, as this approach led to simulation runs in which a chaotic pattern of disconnected road segments emerged.

Yeh and Li's models

Anthony Gar-On Yeh and Xia Li have developed a CA model for simulating rapid urbanization in the Pearl River Delta region of China. The model is designed, substantively, for use in evaluating questions about the sustainability of urban development in that area, particularly in relation to urban encroachment on agricultural and resource land (Li & Yeh 2000, 2002; Yeh & Li 2000, 2001). Recent applications have also seen the model used for testing scenarios for the promotion of compact development (Yeh & Li 2002).

The models are formulated to simulate the conversion of agricultural land to urban uses. An array of constraint variables is used. For the most part, constraints are designed to enable exploration of policies to control urbanization in sensitive areas. The model works in much the same way as the other general urban growth models described in this section. Population projections establish quotas of cells for development. Urbanization is then assigned probabilistically to cells based on a set of determining influences. These include roads and accessibility, agricultural suitability, suitability for urban uses, soil information, slope details, water quality, and development intensity.

The RIKS models

Roger White, Guy Engelen, and colleagues at the Maastricht: Research Institute for Knowledge Systems (RIKS) have developed an elaborate CA model that simulates urban growth (among other things) through land-use transition. The models have been applied to a variety of cities and places, including the Netherlands, Dublin, Lucia, and Cincinnati (Engelen *et al.* 2002; Engelen *et al.* 1995; White 1998; White & Engelen 1993, 1994, 1997, 2000; White *et al.* 1997). In addition, other CA models have been developed, based on the RIKS rule-set (Arai & Akiyama 2004; Bäck *et al.* 1996).

Potential for land-use transition is determined probabilistically using a local distance-decay rule, applied within varying neighborhood filters. Early versions saw transition governed by exogenously-defined quotas for land-use intensity (White & Engelen 1993). Recent versions render transition decisions based on a series of exogenous models that cover rules for the natural environment, demographics, and the economy. Transition is also subject to a series of weights and constraints.

The RIKS models include a number of sprawl-relevant features. Open space is introduced by means of 'fixed' and 'functional' states; 'fixed' cells are excluded from transition, but nonetheless may still factor into the transition potential for neighboring 'functional' cells. Accessibility features through the introduction of distance-decay variables that weight the influence of neighborhood input relatively favorably for sites close to a target cell, compared to those at a distance. Like most automata models, the RIKS simulations are highly dynamic. The connection to exogenous models also allows for the derivation of sprawl-relevant outputs, including loss of resource land. The models have been widely used in exploring public policy scenarios (Engelen *et al.* 2002), although in Europe, rather than the United States.

SLEUTH

Keith Clarke and colleagues at the University of California, Santa Barbara and the United States Geological Survey have developed the SLEUTH model—a comprehensive and generalizable urban growth simulation tool. SLEUTH has seen application to a wide variety of cities, both in the United States (Santa Barbara, Baltimore, and the San Francisco Bay Area) (Candau *et al.* 2000; Clarke 1997; Clarke & Gaydos 1998; Clarke *et al.* 1997; Goldstein *et al.* 2004; Herold 2002; Herold & Clarke 2002) and abroad (Lisbon and Porto) (Silva & Clarke 2002).

Urban growth in SLEUTH is specified via an overall growth rate, which may adapt through self-modification as a simulation evolves. That growth is assigned, geographically, within the model using five transition rules. A 'diffusion' rule is used to disperse growth. The 'breed' rule generates growth spontaneously by agglomeration. A 'spread' rule regulates growth by outward expansion. The 'slope' rule introduces resistance, curtailing growth in areas beyond a slope threshold. Finally, a 'road gravity' rule is employed to mimic growth related to transport infrastructure.

Several of these rules are germane in the context of sprawl. The 'diffusion' rule is quite similar to the notion of scattering of development associated with suburban sprawl. The 'road gravity' rule could be likened to accessibility in its ability to influence urban development. The SLEUTH model also features excluded states, including open spaces and resource land. Both of these are characteristics of relevance to sprawl.

Automata models of polycentricity

A handful of CA models deal with urbanization as a polycentric process. These simulations go some way toward explaining the forces that give rise to sprawl, at least in terms of the formation of sub-centers outside dominant urban cores.

As part of their work exploring self-organization and spatial self-organization in an urban economic and urban growth context, Krugman and Fujita have developed automata models focused on polycentricity (see Fujita et al. 2001, Krugman, 1996 for an overview). Krugman has developed automata models dealing with polycentricity in an edge city context. This is an innovative approach, starting with very general notions of urbanization that date back to Von Thunen's work, which Krugman and Fujita recognize as incorporating spatial organization, but not *self*-organization. The focus in their models is on the emergence of structure from the "internal logic" of the system. Their automata models are based on growth distribution, balancing centrifugal and centripetal forces—dispersal—of clustering. After experimentation, they have found that their models can generate realistic urban evolution. Following a smooth initial distribution of urban centers (business centers), the distribution of urbanization begins to undulate in a rough pattern. A number of small, independent centers then emerge in proximity to each other, before organizing into large dominant center of urban activity. Moreover, the simulated city carries complexity signatures that resemble those of cities in the real world.

Fulong Wu has developed a model for explaining various scenarios relating to the formation of polycentric patterns of urban growth (Wu 1998). Wu used four different rule-sets. Development of land for urban uses is determined by distance from a central core, the intensity of existing development in a target cell's neighborhood, population density, or a combination of development intensity and population density.

Each of these rules has relevance in terms of our sprawl discussion. Distance to a central core has similarities with the notion of overall accessibility within a city-system. Intensity of development is quite close to the idea of local accessibility. Interestingly, Wu found that only combinations of development intensity and population density rules led to polycentric patterns of urbanization in his simulations.

Yeh and Li also developed models of polycentricity, in relation to the exploration of compact growth scenarios (Yeh & Li 2002). Several sprawl-related characteristics feature in the model as state variables. Accessibility is specified in relation to distance to a main urban center, but also as proximity to nearer sub-centers. Weights are assigned to those accessibility variables to favor one form of accessibility over the other, thereby encouraging monocentric or polycentric growth. Density characteristics also feature in the model. Yeh and Li also experiment with different density scenarios for growth, ranging from high density and a rapid density decline from the CBD, to low density and a slow density decline.

Fringe urbanization automata

General growth automata models simulate urbanization as a bottom-up process governed largely by land-use transition. Some of those models also treat growth in terms of general suburbanization forces (the 'spread' function in the SLEUTH model is one example). Automata models of polycentric urbanization represent a step closer toward models of sprawl that resemble the phenomena discussed in Chapter 6. However, there is also a handful of automata models designed explicitly for exploring the forces of urbanization on the suburban fringe.

Bell and colleagues at Adelaide University's Key Centre have developed a CA model of fringe urbanization in South Australia (Bell *et al.* 1999). Their CA model is nested within a series of forecasting models, and handles the assignment of growth (demand) to small area geographies (of one square kilometer in size) in the simulated city.

The focus of the model is on determining the probability of a particular site being converted from greenfield to urban status. Probability is specified based on a number of attributes. These include institutional factors, accessibility, and contiguity.

Each of these determinants is relevant in the context of sprawl. The explanatory variables are flexible—the model is designed as a PSS and variables can be altered by users. Nevertheless, the usual set of variables resembles those mentioned in Chapter 6, in relation to the characteristics of sprawl. Institutional factors include zoning. Accessibility factors incorporate accessibility to a range of activities such as schools and shops as well as to the CBD. The contiguity variable is particularly interesting. New development is designed to occur with greater propensity *close* to existing areas, rather than in isolated locations. The opposite is generally true in the context of North American cities, where leap-frogging is more likely to occur than cohesion.

Wu developed a model of fringe urbanization in the Tianhe area in suburban Guangzhou, China. The paper describing the model reports mostly on the calibration procedures used in its development (Wu 2002). Nevertheless, the model deals only with fringe urbanization, and is therefore quite relevant to sprawl.

Model rules are focused on explaining the probability of land conversion from non-urban to urban uses. Development is determined based on a set of variables, including accessibility to a central core. Interestingly, the model simulates new suburban development only, by constraining urbanization on already-developed sites. In other words, it prohibits re-development. The model does not deal with density *per se*, but development probabilities are weighted by a distance-decay parameter. This parameter is used to yield higher potential to development in areas on the urban boundary, with opportunity declining with distance from the fringe.

The complex systems and environmental science groups at the University of Michigan have developed prototype models for exploring the ecological effects of sprawl in abstract cities (Rand *et al.* 2002). Their model is agent-based and deals explicitly with sprawl, simulating how the preferences individual relocating households make lead to sprawl-like urban patterns.

Agents in the model are programmed with two behavior rules: and one that expresses their preference for density (the presence of other agent households) and another that expresses preference for 'natural beauty' (a dummy variable that also incorporates the intensity of settled agents in a neighborhood). The authors used the model to explore some preliminary scenarios about sprawl, concluding that both factors determine the level of sprawl in a simulation (measured in terms of the clustering of agents into cohesive or spatially-separated units), independently, and with combined influence. Initial settlement is attracted to the center of a simulated city, coaxed by the situating of 'service centers' in that location. Greater preference for natural beauty leads to the relocation of agents on the very fringe of the simulated urban space, after initial settlement of the center was underway. An increase in preferences for density leads to less sprawl; reduction in that preference leads to random clustering. However, the density rule was found to have less of an impact, relatively, compared to the beauty rule. Moreover, when the two rules were combined, the collective impact was much larger than that observed with the use of single rules alone.

In other applications, green belts were introduced to the simulation, and their effect on simulation dynamics was examined in a preliminary manner (Brown *et al.* 2002).

There are also several models of fringe urbanization in the land cover literature, but these tend to simulate deforestation and transition between canopy types as a result of urbanization, rather than sprawl-like development. A thorough review of those models is available in Parker *et al* (2003).

Conclusions

Chapter 6 provided a review of the literature on sprawl: its characteristics, consequences, and potential causes. This chapter has reviewed the literature on urban models as it relates to sprawl, exploring ways in which sprawl and elements of the phenomenon have been modeled in the past.

This chapter discussed the requirements for a robust sprawl model. General requirements for dynamics, space and geography, and inter-dependency between sub-systems were identified. The set of key variables that might feature in a sprawl model, representative of the characteristics of sprawl, were explored. Some methodological considerations in relation to representation of the proposed causes of sprawl were also discussed.

The literature on urban simulation was appraised as it relates to sprawl. Population density models were reviewed, as were some relevant PSS. Particular attention was paid to urban automata models: general models of urbanization, models of polycentricity, and fringe urbanization simulations.

Two sets of conclusions may be drawn from the discussion in this chapter: those relating to the use of simulation to study sprawl and urban systems more generally, and methodological implications.

Implications for urban studies

Very few (if any) models have been developed to study suburban sprawl *explicitly*. This is surprising considering the currency of the topic. For the most part, the literature is focused on simulation of general urbanization and this is particularly true of the automata literature. It is unfair to critique the literature in terms of sprawl when it is not intended for that application. Nevertheless, the discussion in this chapter raises some issues relating to the general treatment of urbanization in the literature. The automata field is in its infancy; research is focused on adapting methodology—originally developed for uses in mathematics, the computing sciences, the physical sciences, and non-spatial social sciences—for geographic uses. However, there are some noteworthy omissions. Decentralizing forces do not generally feature—the mechanisms that actually lead to suburbanization (and sprawl). Most models do not differentiate between urbanization in core urban areas and that on the urban periphery. The models of fringe urbanization discussed in earlier sections deal with areas of the edge of an urban mass, but do not treat urbanization processes any differently from those in the urban core. Conversion from agricultural to urban uses is mostly handled like any other general land-use transition. Also, urban automata models are generally weak in their representation of links between micro- and macro-phenomena. Automata models have the *ability* to facilitate micro-macro linkages, by design. However, relatively little attention has been devoted to specification of transition rules to account for the scaling of processes. Models are specified, wound up, and let go, but the *path* between local and global scales is not always traced out in the actual mechanics of the model. This is true in spite of growing appreciation for the importance of cross-scale forces in the evolution of urban phenomena. The prevalence of 'postmodern' (or 'post-Fordist') approaches in planning literature, with its emphasis on the 'grassroots' of urban phenomena, is one such example (see Soja 1995). Generally speaking, urban automata models are not as well-equipped to answer policy-relevant or theory-relevant questions as PSS. In fact, PSS like UrbanSim are getting more like automata tools in their simulation capacity, while retaining their policy potency. (And not the other way around.) Again, this is somewhat unfair; automata modeling is relatively young as a field and research is focused on other priorities (Torrens 2002).

Implications for simulation

The discussion presented in this chapter has methodological implications. The apparent difference between conventional approaches (population density modeling, PSS) and 'new' approaches characterized by automata tools is important in terms of simulating sprawl.

The PSS discussed are mostly based on regression equations (linear regression, hedonic price models, logit models), with results being applied to local-scale geographical structures (polygons, rasters) in GIS. PSS are getting more sophisticated: the 'micro-simulation' trend has introduced an emphasis on disaggregate variables, for example. With the exception of UrbanSim, however, the emphasis is on *accounting for* behavior. The PSS models that were discussed focus mostly on explaining variance in price and the assignment of surplus demand to supply rather than simulating actual decision-making behavior. This limits the sort of questions that can be asked with the models; it also means that many of the factors relevant to the sprawl debate cannot be explored.

The treatment of dynamics in PSS is also problematic. The PSS models discussed in this chapter are longitudinal. They may be specified at fine temporal scales, even at the scale of cycles of residential housing markets or development schedules. But, again, this abstracts from the dynamics of urbanization *processes*. The inability to handle system evolution through parallel and interactive dynamics means that some important properties of system dynamics are being missed; in particular, self-organization and associated path dependence and lock-in, emergence, and innovation.

PSS models are disaggregate and heterogeneous in their treatment of data inputs, but homogenous in terms of representation of actual system *behavior*. The processes driving urbanization are not treated heterogeneously. Interaction is largely treated as if it takes place evenly across the system. Moreover, the models mostly ignore individuals as *process-makers*. Individual-based data points may be fed into the models (e.g., land parcel conditions), but everything delegates to a universal set of regression equations. This differs from the automata approach, where agents and cells may be specified with a few transition rules, but they act independently, locally, and interactively within the system. The failure of PSS to account for interaction in space is very significant; behavior in the model is reduced to an assignment of data points to appropriate locations. This is like the 'hidden hand' approach that was historically popular in economics, with associated assumptions of centralized control (for example, see Krugman 1996; Resnick 1996, 1997). By contrast, automata tools provide the opportunity to 'grow' cities, from the bottom-up. This is an important distinction in terms of sprawl: many of the potential causes outlined in Chapter 6 operate from the ground-up; few are a function of central executives. This is part of the problem: growth management is centralized, and there may be need for individual-level mechanisms to encourage smart growth.

A by-product of this is that there are some important limiting—and unrealistic—assumptions in PSS and models of this nature. Examples include perfect knowledge (access to all available opportunities: available housing, available development options), equilibrium assumptions between demand and supply (UrbanSim is an exception), reliance on artificial mathematical market-clearing mechanisms instead of behavior, and the assumption of rational (price-maximizing or utility-maximizing) behavior.

Criticism may also be leveled at automata models. Many of the automata models discussed in this chapter are 'blob' models (as Catherine Dibble has termed them in some recent conversations we have had!). The focus is on generating realistic patterns of urban growth (urban blobs), but there is comparatively little emphasis on representing the agents of change. This is particularly true of urban growth automata. This is, perhaps, a consequence of the application domain of the models, but is also likely a function of validation techniques and the emphasis on pattern instead or process.

The lack of uptake for agent-based tools is also significant. With the exception of the SLUCE models, each of the simulations discussed was CA-based. The general flexibility of the neighborhood concept may be responsible. The simple mechanism of referencing a target cell to average, maximum, or minimum conditions in a neighborhood filter, with additional distance-decay constraints, is a powerful concept, facilitating simulation of a variety of processes of relevance to urban systems. However, agent-based tools and 'agency' concepts have yet to be explored, popularly, in the simulation of urbanization. (The work of Denise Pumain and Lena Sanders (Sanders *et al.* 1997) stands out, although, this is largely CA-based, with some agent-like interpretation of cells.) The rich potential for creating synthetic urban agent entities has not been fully realized outside of the transport literature (see Torrens 2004a). Of course, the agents of urbanization differ considerably from pedestrian or vehicular agents: they are not constrained by the rules of the road or the physics of highway and sidewalk travel.

What is perhaps most noteworthy, in the context of this thesis, is the relative lack of geography in many urbanization automata models. The models resemble Conway's Game of Life rather than Reynolds' Boids. Many of the models are transition models, relying on conditions in a neighborhood filter to determine the probability of a state change. Transition is the base point for explanation in such models. Often, there may be no mention of the events or processes that led to transition. In the context of sprawl, for example, there are a variety of geography-specific *processes* that cause transition from one land-use to another—movement, migration, relocation, etc. Again, this makes it difficult to simulate some of the factors hypothesized to drive sprawl.

Part of the reason why automata models have these difficulties is that there is not an adequate framework for infusing geography into urban automata models. The Geographic Automata Systems framework outlined in Chapter 5 can help, by providing the basic functionality necessary to describe geographic processes. It also offers advantages in facilitating the specification of agent-like functionality, symbiotically, with cell-like behavior.

The part of the thesis discusses the development of a series of models of suburban sprawl, using the Geographic Automata Systems framework outlined in Chapter 5, and considering the discussion presented in this chapter and attributes of sprawl outlined in Chapter 6. Several aspects of the models borrow from the methodology and approaches mentioned in this chapter, and in many ways the models attempt to remedy the weaknesses raised in this discussion.

Part three: empirical applications

Chapter 8. The spatial distribution of growth in a sprawling system

"Program a map to display frequency of data exchange, every thousand megabytes a single pixel on a very large screen. Manhattan and Atlanta burn solid white. Then they start to pulse, the rate of traffic threatening to overload your system. Your map is about to go nova. Cool it down. Up your scale." (Gibson 1984, p. 57)

Introduction

Part three of the thesis describes development of empirical applications for studying sprawl. Two models are described. The first, discussed in this chapter, models supply-side elements of sprawl, focusing on the roles of growth, scatter and fragmentation, density, development, and transport in sprawl dynamics. The second, discussed in the next chapter, models sprawl from a demand standpoint, focusing on the role of preferences in determining very local-scale dynamics in sprawled residential communities.

The model discussed in this section is designed to test the application of the Geographic Automata Systems (GAS) framework to the simulation of sprawl in a generic, abstract, city-system. The model is used to simulate the spatial distribution of growth, and the subsequent evolution of an urban system over time, with emphasis on the patterns of urbanization (sprawl) that are generated and the pace with which the phenomenon operates. In the simulation, a city-system evolves from initial seed settlements, going through processes of compaction, poly-nucleation, infill, peripheral sprawl, and densification of the central city. The model was designed with careful attention to the attributes of sprawl discussed in Chapter 6, and drawing inspiration from the body of literature mentioned in Chapter 7, and the issues raised in that chapter.

As a model of sprawl, the model described in this chapter is designed to simulate the spatial distribution of development, settlement, and its density, dynamically as it evolves over time. The processes are simulated as a function of interactions between humans and their environment. The mechanisms driving transition in the model are specified with consideration of the proposed causes of sprawl outlined in Chapter 6. The model contains mechanisms for poly-nucleation, scattering and leap-frogging, transport-influenced growth, decentralization, and downtown decline. Individual simulation runs are also evaluated in their capacity to generate realistic patterns of sprawl, and their ability to evolve in a realistic manner.

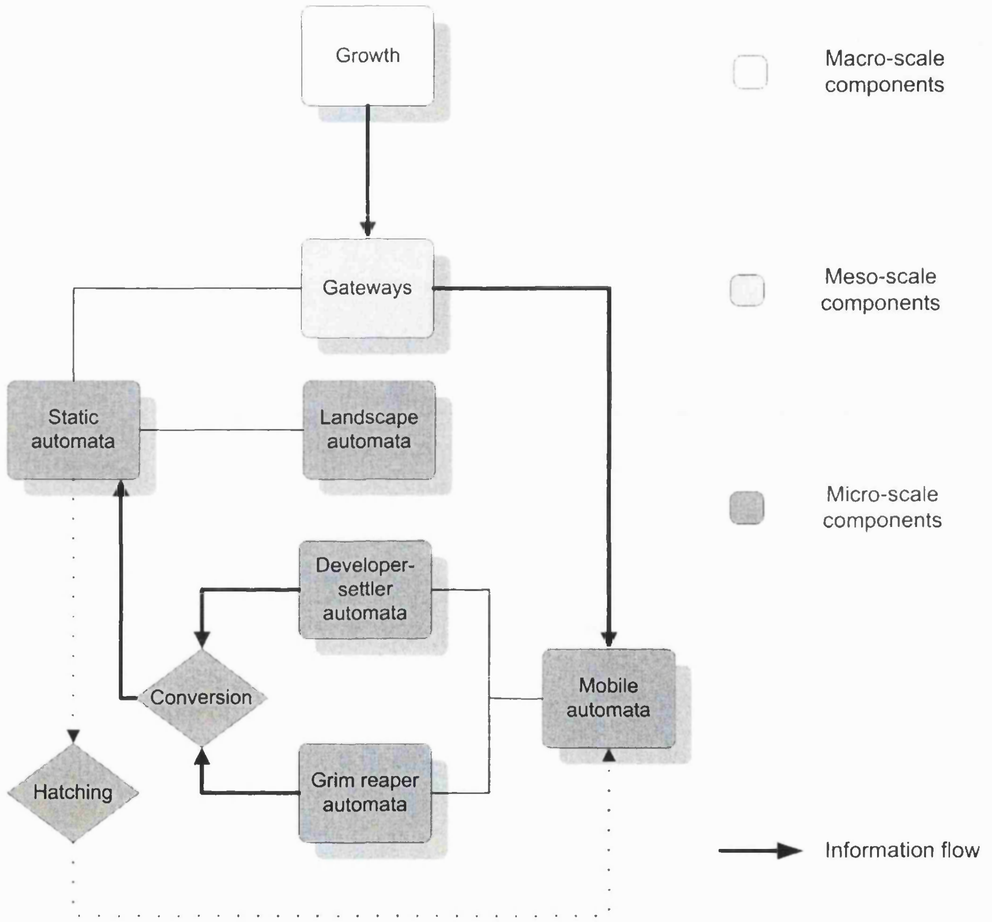


Figure 31. Overview of the spatial distribution model

Model specification

The model has been built using the GAS framework outlined in Chapter 5, and programmed in StarLogoT 2001 (Center for Connected Learning and Computer-

Based Modeling 2001). A simplified diagram of model components is outlined in Figure 31. Recalling the GAS discussion, a GAS is composed of interacting Geographic Automata (GA). GA retain the attributes of general automata, CA, and MAS discussed in Chapter 4, but are designed with additional, geography-specific, functionality to help animate them in space-specific contexts.

$$G \sim (K; S, T_S; L, M_L; R, N_R) \qquad \text{Eq xxxix}$$

As outlined above, a GAS (G) is defined with reference to a typology of GA entities (K) that may be fixed or non-fixed in space. Individual GA are characterized with state variable (S), and transition between state variables is determined by a rule-set (T_S). Non-fixed GA are spatially animate and their location at any point in the evolution of a simulated system is expressed via location conventions (L); their motion is driven by a dedicated movement rule-set (M_L). Adjacency and proximity relationships between automata are expressed via neighborhoods (R) around GA, and a neighbor rule-set (N_R) determines changes in those relationships.

Geographic Automata typologies

Entities are divided by typology in the model, into GA that are fixed in space and those that are not fixed. Fixed automata are used to represent infrastructure elements. The (geographical) agents of change (sprawl) are represented as non-fixed GA in the model. Mobile GA are designed to represent developers responsible for converting land to urban uses, and settlers responsible for populating that infrastructure at particular densities of settlement. Fixed and non-fixed GA are complementary in the model. Fixed infrastructure GA serve as a 'container' for mobile developer-settler GA.

The state variables, transition rules, location conventions and movement rules, neighborhood and neighborhood rules that specify GA of fixed and non-fixed types are designed to mimic infrastructure, developer, and settler characteristics and behaviors as they refer to attributes of sprawl discussed in Chapter 6.

State variables

There are two types of mobile GA in the model: developer-settler and settled GA. General state variables are used to simply assign these designations to those GA, and GA of each type behave accordingly in the model.

It is not feasible to represent all of the sprawl characteristics from Chapter 6, but infrastructure GA are specified with state descriptors that mimic a subset of those sprawl characteristics—land-use and density. Initially, fixed GA are denoted with a 'landscape' land-use, used to indicate fixed GA that are not in urban use. This is equivalent to agricultural or natural resource land. In addition, fixed GA are characterized with a binary state variable: 'developable' (Table 6). This is used to denote sites in the simulated city that are open to development or may be exempt from the urbanization process. The state variable is therefore roughly equivalent to the functionality or exclusion variables used in urban CA models (Clarke & Gaydos 1998; White & Engelen 2000; Yeh & Li 2001). If a 'developable' GA becomes developed in subsequent time-steps, the land-use state is transitioned to a 'developed' condition; it is converted from non-urban to urban uses. Under certain conditions, a 'developed' cell may come under redevelopment forces; in these instances, a 'vacant' land-use state is assigned to the GA to denote land that has previously been in urban use but is now vacant ahead of possible re-development. 'Vacant' GA do not transition to 'landscape' conditions; once converted to urban uses, fixed GA may not return to non-urban uses. This is synonymous with the notion of 'brown-fields' sites in cities. Fixed GA have an additional state variable for assigning *settlement density* to infrastructure GA. The density of settlement is simply specified as the number of settled GA occupying an infrastructure GA at a particular point in time.

Table 6. Model state variables.

Attributes	*Data type*	*State values*
Fixed GA (cells)		
Developable?	Binary	Yes, no
Developed?	Binary	Yes, no
Gateway?	Binary	Yes, no
Land-use	Categorical	Urban, non-urban, vacant
Density	Continuous	
Non-fixed GA (agents)		
Active?	Binary	Yes, no
Function	Categorical	Developer-settler, grim reaper
Population	Continuous	

State transition rules

Behavioral realism was one of the main goals of the modeling exercise. Transition between the states mentioned in the last section is largely governed by the *interaction* between mobile GA and fixed GA, which is how change takes place in *real* urban systems. Developer-settler GA are animated, spatially, by a range of movement rules, and these largely govern how they interact with the infrastructure and landscape in which they operate; movement rules account for many of the causes of sprawl identified in Chapter 6.

Developer-settler GA instantiate state transition *directly* in the fixed GA that they encounter. The *presence* of a mobile GA of a given state determines much of the transition in non-fixed GA. If a developer-settler GA occupies a ('developable') 'landscape' GA, the state of that fixed GA is changed from 'landscape' to 'urban'. On the other hand, if a mobile grim-reaper GA occupies a fixed 'urban' GA, its land-use in switched to 'vacant'.

The density state of fixed 'urban' GA is determined in a similar manner. When a developer-settler GA occupies a fixed 'landscape' GA, it switches its land-use state to an 'urban' condition, but also 'settles' that GA by depositing a number of 'settlers' in the GA. If the fixed GA in question is already in a state of 'urban' use, the developer-settler GA then *adds* to the value of settlers in the site, thereby increasing its population density. Similarly, if a mobile developer-settler GA occupies a fixed 'vacant' GA, it can convert it to an 'urban' land-use and settle it with a given density. Non-developable fixed GA are immune to these transitions.

On occasions, a developer-settler GA may also designate a fixed GA as a 'gateway'. This is roughly equivalent to the introduction of urban sub-centers—edge cities—within the urban system.

An additional state transition rule is used for fixed GA. At each time-step in a simulation's evolution, a proportional increase or decrease in the value of fixed GA's 'density' variable value is initiated. This is achieved by taking the density value in a target GA and adding or subtracting a given percentage of that figure to GA within the target GA's neighborhood. This is generally set randomly. The rule is used to mimic very local-scale diffusion of population or decline.

A single state transition rule is used for mobile GA, to determine whether they are 'alive' or 'dead' in a given time-step. This is tied to the movement of the GA. Once a developer-settler GA initiates a transition in a fixed GA (i.e., once it deposits its population), it 'dies' and is removed from the simulation.

Geo-referencing conventions

Two geo-referencing conventions are used to track the movement of GA within a simulation. One is direct, the other is indirect. GA have a set of coordinates that register their actual position in the simulated city. These coordinates are represented as simple X and Y values that indicate the position of the GA raster in the overall raster matrix. Mobile GA are also described with X and Y coordinates, but have an additional indirect location reference, denoting the gateway from which they originated.

Movement rules

Movement rules characterize the core behavior in the model—they introduce geographical processes into a simulation run. Movement is specified, essentially, as a constrained and targeted random walk. The end result of movement is a state transition, as explained previously. In the last chapter, lack of process-based mechanisms was often cited as a criticism of the reviewed models. By introducing movement as a pre-cursor to state transition, geography-specific behaviors and processes can be used in a simulation, thereby enabling exploration into the role of spatial processes in the dynamics of sprawl.

A set of movement rules is used to spatially animate GA. They resemble the rules used in the SLEUTH model (Clarke & Gaydos 1998; Clarke *et al.* 1997), but there are differences: these rules are designed with explicit consideration of the causes of sprawl outlined in Chapter 6; they are also *agent-based* rather than CA-based in nature. GA actually move, rather than diffusing state information from static positions.

Potential causes of sprawl were reviewed in chapter 6. Among a multitude of causes, several were determined to be significant in driving suburban sprawl: urban evolution; population growth; expression of preferences for suburban living; decentralization of economic activity; transport and telecommunications; developer behavior; planning and public policy; and political fragmentation. The movement rules were designed with these causes as a consideration. As mentioned previously, growth is implied at the onset of a movement event in the model. The goal of the model, then, is to distribute growth (or decline) over a simulated landscape, on the basis of a set of spatial heuristics. How do these rules relate to the causes discussed in chapter 6? As will be discussed toward the end of the chapter, the model does a reasonably good job of simulating urban evolution through various growth stages. Population growth is handled via the growth rate mechanism. Decentralization of economic activity is implied in so much as the model, at the movement step, is designed to distribute growth in a decentralized manner, albeit around intra-urban cores and with the possibility of quite localized growth (the immediate and nearby movement rules). Transport and telecommunications are represented by means of a road-like movement rule. Expressed preference for suburban living is represented by means of a leapfrog rule. Development is implied throughout; the agent entities in the model are characterized as developer-settlers, and it is assumed that the movement rules represent the physical manifestation of developer and settler agents in an urban setting.

Mobile developer-settler GA exercise either one or many movement rules when they exit a gateway. This is where much of the experimentation in the model comes into play, in emphasizing certain movement functions over others or looking at the isolated and relative influence of individual movement rules. The rules can be weighted, probabilistically, so that some rules are more likely to be executed than others. Mobile GA can also be set to exercise rules in a particular sequence.

Immediate movement: The immediate movement rule mimics initial development processes, whereby a site is settled very locally. Under this rule, a developer-settler GA wanders randomly within a very confined neighborhood consisting only of the fixed GA immediately surrounding a target site in a Moore configuration (Figure 32). At each fixed GA that is traversed, settlement takes place. Generally, this results in very compact development (settlement) in a confined radius around a target site.

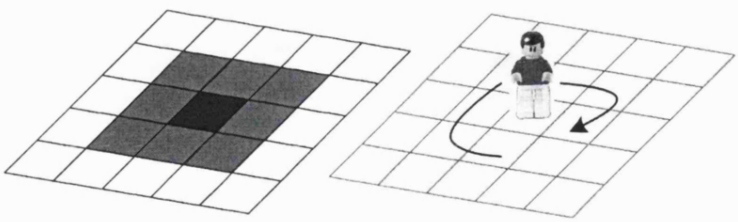


Figure 32. Immediate movement.

Nearby movement: The nearby movement rule is similar in specification, except the neighborhood window for movement is much larger in size (Figure 33). This rule is used to simulate larger development projects where several developments are constructed in close proximity. Generally, the rule yields *clusters* of settlement, equivalent to New Urbanist (Calthorpe *et al.* 2001) or transit-oriented village (Cervero 1998) types of patterns.

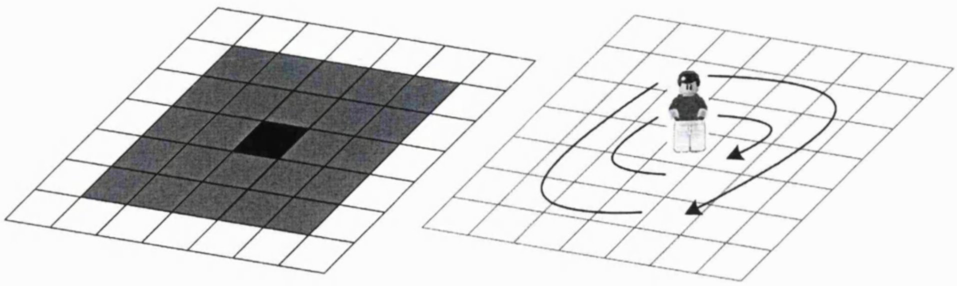


Figure 33. Nearby movement.

Irregular movement: The irregular movement rule is used to simulate situations in which development is constrained and must proceed in an irregular fashion, e.g., because of natural boundaries such as mountains, rivers, wetlands, etc., or because of administrative boundaries (Figure 34).

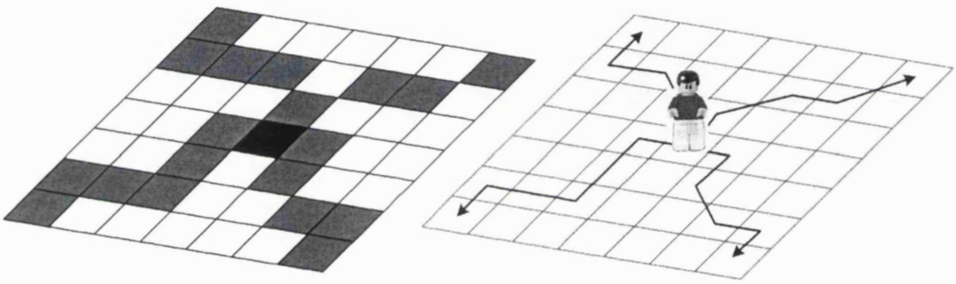


Figure 34. Irregular movement.

Road-like movement: The road-like movement rule is used to mimic road-building. Previous automata-based models of urban growth have introduced road development as an accretive process—roads 'grow', sequentially, by diffusion-limited aggregation (Xie 1996). However, roads do not develop in a piecemeal fashion in the real world (see Ward, Murray *et al.* 2000 for a discussion of the problems they had growing roads in their models); rather, they are constructed as links, but only open once completed. The approach taken in this model is quite different. Roads are developed as nodes, first, and then those nodes are connected by strips of development, indicating transport-oriented growth flanking road infrastructure. Developer-settler GA move by means of one of the other movement rules, but lay down nodes instead of population (Figure 35).

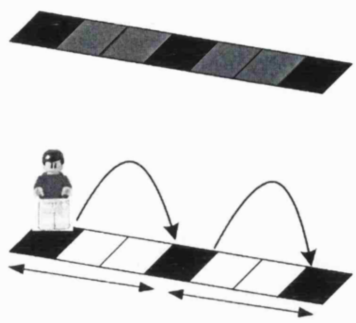


Figure 35. Road-like movement.

Leap-frog movement: Developer-settler GA may also move by leap-frogging. Under this rule, GA move in 'jumps', settling the fixed GA they encounter on the termination of each 'hop' (Figure 36). This mimics the leap-frog development patterns associated with sprawl that were discussed in Chapter 6.

In addition, rules may be combined—a developer-settler GA can exercise rules in isolation or can execute a sequence of rules before terminating its movement. For example, after moving by leap-frog, a GA might initiate either an immediate or nearby movement. Depending on which rule followed the leap-frog, the resulting pattern would be a sprinkling of isolated settlements or more polycentric forms consisting of adjacent clusters that may fill-in through diffusion.

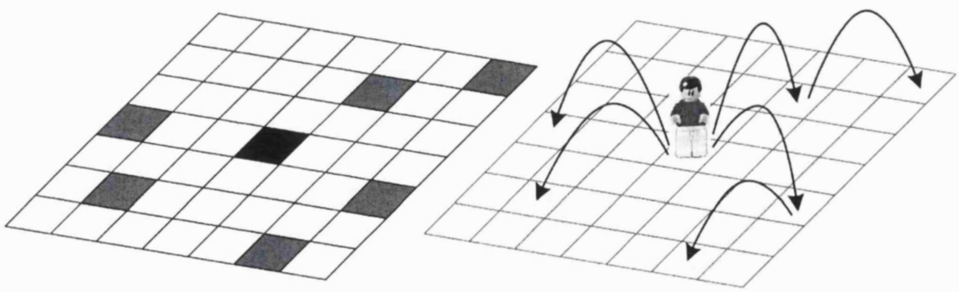


Figure 36. Leap-frog movement.

Neighborhoods

Two neighborhood windows are used in the simulation, both specified in a Moore configuration. The first consists of a target cell and its eight adjacent cells; the second is an extended Moore neighborhood of 24 cells (Figure 32, Figure 33). However, the use of movement rules allows for the introduction of action-at-a-distance.

Constraints

A variety of constraints are introduced to the model to confine simulation runs within specified bounds. As was mentioned, fixed GA can be coded as either 'developable' or 'non-developable', allowing for certain areas of the simulation to be withheld from transition. The specification of gateway GA introduces a spatial constraint, tying state transition to certain core sites in the simulation. The hierarchy of land-use transition adds a further constraint that ensures realistic transition of fixed GA between uses.

The overall rate of evolution in the model is specified using a growth rate and it acts as a constraint on the rate of development and settlement simulated by the model. This is synonymous with the impact that population growth has on suburban sprawl, as discussed in Chapter 6. Cities with rapid growth will likely expand—and sprawl—at a greater rate than might occur under slower rates of growth. As discussed in Chapter 6, growth rates are usually specified as quotas for state transition. However, this model follows the approach adopted in the PSS described in Chapter 7—an actual volume of growth is introduced to the model at a given time step. In the PSS examples, growth is assigned locally to rasters using a market-clearing mechanism. A different approach is used here—a volume of developer-settler GA are introduced to the simulation and they go on to 'assign' growth, *spatially*, within the system using theoretically-informed movement rules. If the growth rate is increased, the number of active developer-settler GA in the system grows proportionally. Developer-settler GA enter the simulation via gateways and radiate out through the simulated system from those sites with declining probability as distance increases. This has the effect of concentrating initial development and settlement in the core areas that are initially established in the simulation. Gateways may be defined *a priori* or can pop-up spontaneously as a simulation evolves.

As mentioned, on some occasions developer-settler GA will designate a fixed GA as a gateway. When this happens, that fixed GA automatically spawns a random number of developer-settler GA. This is used to mimic the capacity for sub-centers within an urban system to act a polycentric core in their own right. It also allows for a certain volume of 'endogenous' growth to be introduced to the system from within.

Grim-reaper GA may also be introduced, initiating decline as they interact with fixed GA in the simulated city. This is used to simulate spontaneous urban decline. In this sense, the model allows for a certain amount of endogenous 'birth' and 'death', roughly equivalent to internal demographics in a city-system.

Simulation experiments

Using the model, simulations were built based on two scenarios for sprawling urbanization within an abstract city-system. Both simulations make an attempt to evolve a city-system in a realistic fashion, with emphasis on the patterns generated by the simulation and the rate of simulated urbanization. The following sections describe the two models. Following this, a simulation applied to a *real* city system is described, before a validation exercise, focused on the evaluation of landscape configuration, is discussed.

In these models, it is assumed that the rate of growth is known *a priori.* (In the Midwestern simulation, growth rates are based on population data from the United States Census.) A variety of methodologies exist for simulating growth, including cohort-survival models and input-output models. These are largely beyond the domain of this thesis, but are detailed in Isard, *et al* (1998).

Abstract simulation one: general growth

In the first example, the model is used to build a simulation in which the evolution of a dominant central city is simulated in the context of a larger city-system with two additional, competing, urban centers (Figure 37).

The transition rules described in earlier sections are active in the simulation with roughly equal propensities for execution.

The simulation is programmed with initial seed conditions that introduce gateway sites in five locations: the center of the lattice, two sites in almost immediate proximity, and two other gateways on the right and left areas of the lattice space (Figure 37).

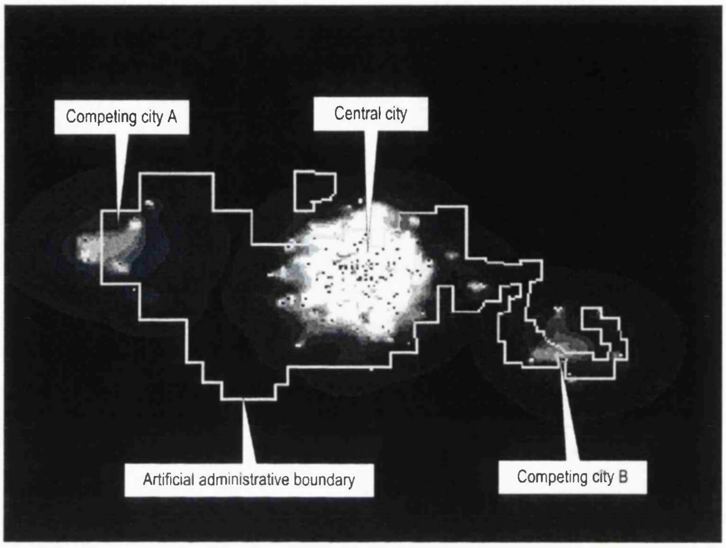
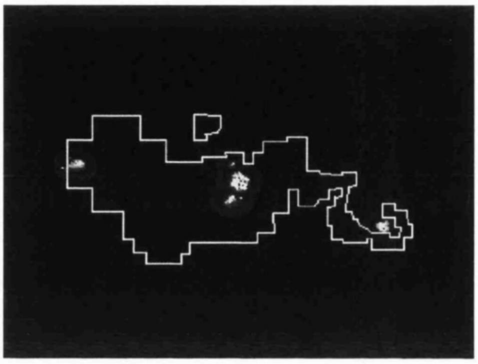


Figure 37. The central city and its two competing neighbors (t = 313)

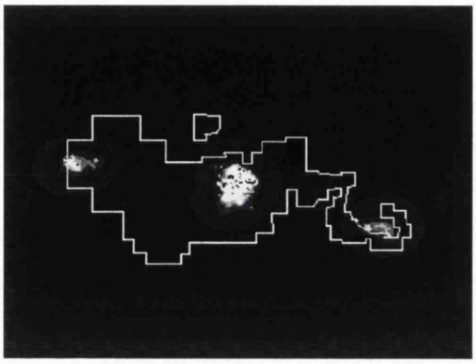
The ability for the emerging cities to compete for space as they sprawl is specified in two ways. First, the central city is afforded an advantage from the start of the simulation by virtue of the introduction of two adjacent gateway sites. Second, the growth rates of the cities are treated differently, thereby influencing the temporal evolution of the urban system as well as its spatial development. The supply of growth to competing cities A and B is cut off roughly 75% of the way through a simulation run, mimicking conditions whereby the critical mass of a dominant central city begins to draw incoming migration and activity away from cities with comparatively less attraction. In the simulation, this occurs when the hinterlands (suburbs) of the competing cities meet those of the central city (Figure 38).

At this point, exogenously-derived growth (in-migration) ceases in the peripheral cities and only endogenous growth continues in those cities.

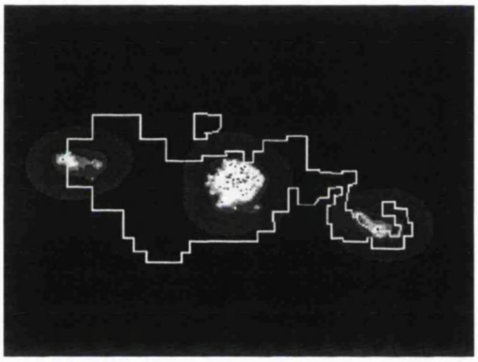
The model does a good job of generating realistic urbanization. The patterns of growth generated in the simulation are synonymous with those that would be expected in a real city-system. In addition, the timing of evolution of the system appears to be realistic (Figure 38).



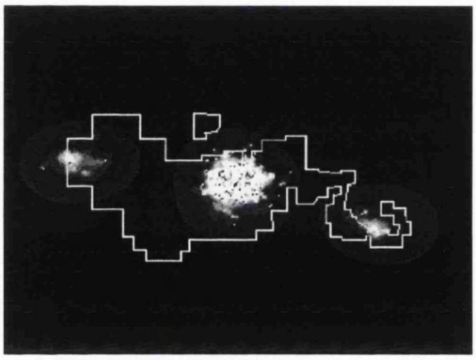
Iteration t = 50



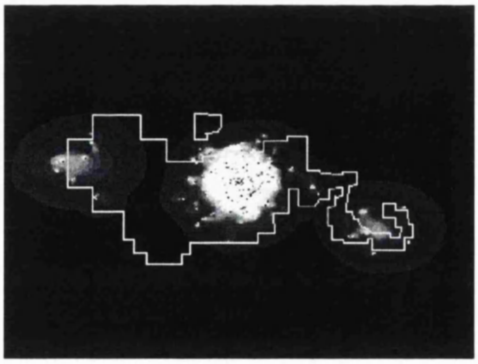
Iteration t = 100



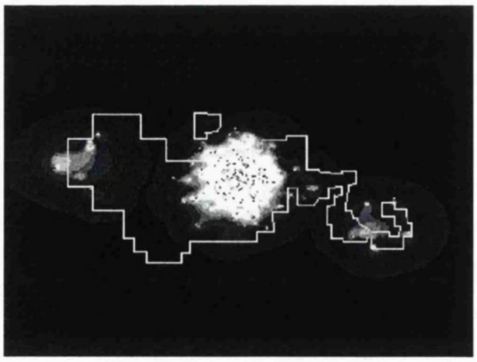
Iteration t =150



Iteration t = 200



Iteration t = 250



Iteration t = 300

Figure 38. The evolution of abstract simulation one

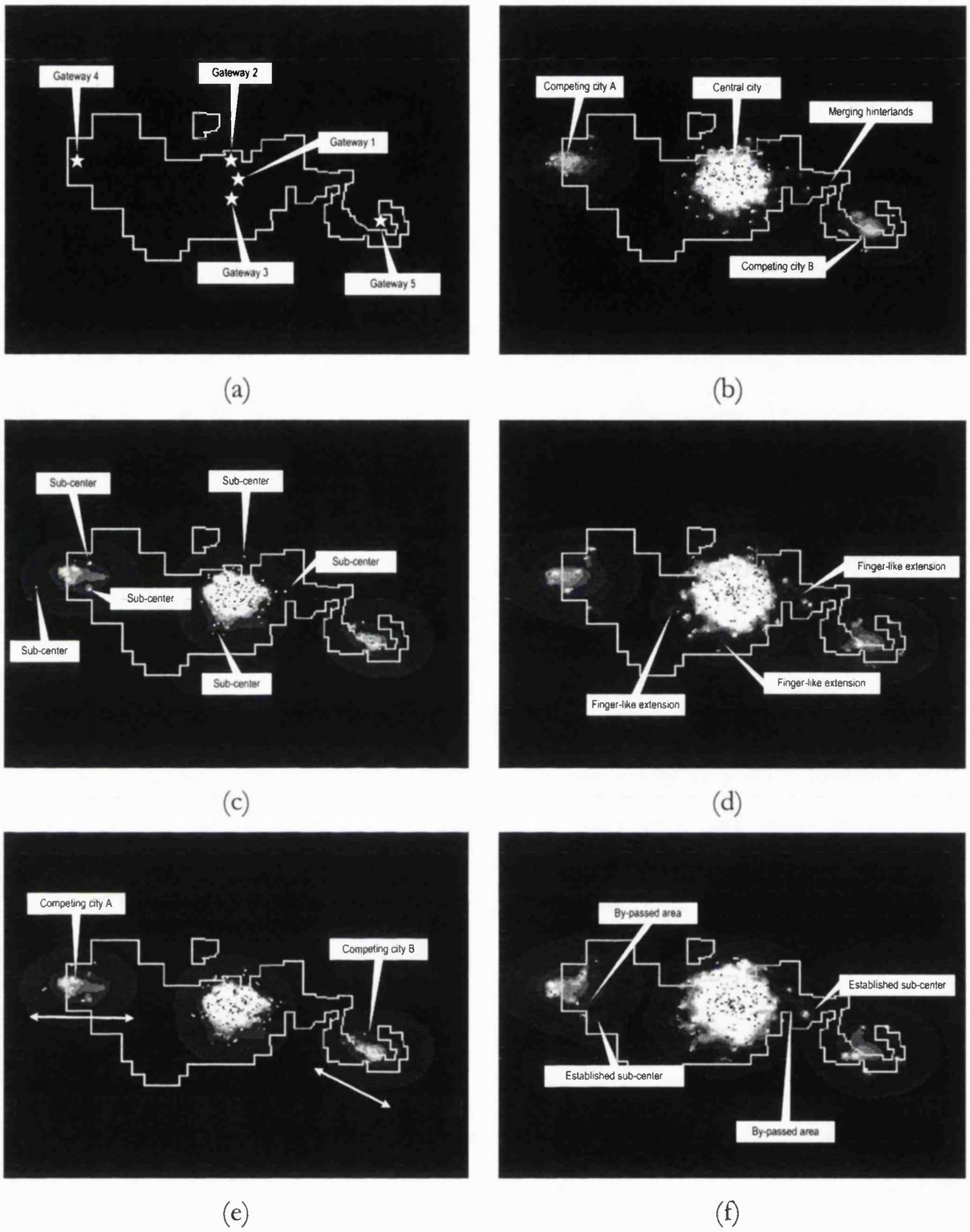


Figure 39. Noticeable features in abstract simulation one

(a) Gateway sites (t = 0). (b) Merging hinterlands (t=222). (c) The formation of sub-centers (t = 186). (d) Corridors of development (t = 260). (e) Linear development (t = 251). (f) Well-established sub-centers, with by-passed interstitial areas (t = 291).

The pattern of urbanization generated by the simulation is commensurate with real-world conditions at a region-wide level of observation. The three cities begin their early evolution as compact cities: dense mono-centric masses with a surrounding lower-density suburban hinterland. As the density of settlement in these centers grows, the expanse of the suburban hinterland extends further in the simulated space, and at an increasingly rapid rate. (The diffusion rule actively disperses a greater volume of settlement as the mass of settlement in the system grows.)

At t =186 (roughly 50% of the way through the simulation run) the effect of the leap-frog rule becomes more pronounced; the urban mass has grown, spawning a greater number of sub-centers on the periphery of the cities. Around this time, the road rule also begins to generate interesting patterns—'fingers' of dense development begin to emerge, manifesting as corridors of growth extending from the main urban mass (Figure 38).

Competing cities A and B actually begin to develop a linear-like development pattern, succumbing to path-dependence because of some initial road-like development (Figure 38).

By t = 227, some suburban sub-centers have begun to evolve as centers in their own right, and the overall structure of the central city becomes largely irregular, with pockets of lower-density settlement that have been by-passed by the urbanization process evident within the evolving city mass (Figure 38).

At t = 150, the hinterlands of the central city and competing city B have sprawled to such an extent that the two urban masses begin to merge (Figure 39). At this point, the supply of growth to competing cities A and B is *stopped.* The downtown areas of the competing cities rapidly begin to decline in density, as growth diffuses throughout the system without a replenishing supply to the gateways of competing cities A and B. By t = 200, the peripheral cities have become largely dispersed, with the remnants of formerly-dominant central seed areas barely visible (Figure 39).

Overall, the city-system sprawls dramatically, while maintaining a realistic pattern of regional-scale urbanization. It is particularly noteworthy that the spatial extent of the entire city-system evolves to a condition whereby the low-density suburbs cover roughly the same area as the denser central cores. Of course, the low-density of that sprawled area means that those sections of the simulated city house a minority of the population.

Abstract simulation two: polycentric growth

In the second example, a second simulation (simulation two) is devised in much the same way as the last example, with identical growth rates and seed conditions, and the termination of growth at a point in the evolution of the simulation. However, in simulation two, the weighting of rules is adjusted to encourage more poly-centric development. The rules are specified with greater propensity for the formation of peripheral clusters.

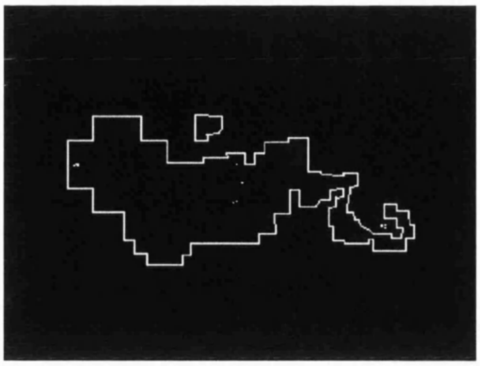
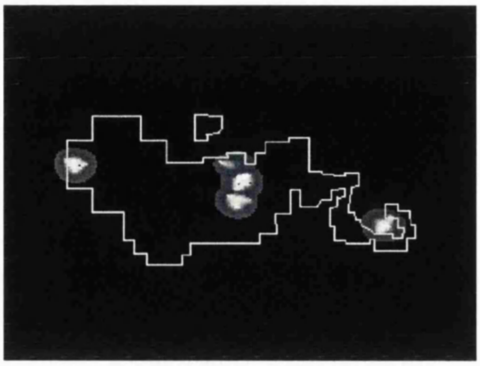
This simulation, essentially, operates under a smart growth regime. Growth is accommodated, but focused in a polycentric fashion. This is achieved using combinations of leap-frog, road, and irregular movement rules as part of a combined sequence that terminates in a nearby movement rule. The likelihood of these clusters being designated as gateway sites is also increased. This sequence is used alongside normal execution of the other rules in isolation. The sequence creates a large number of dense peripheral clusters—edge cities—as a simulation evolves (Figure 40, Figure 41).

The city-system evolves at a much faster rate, due to internal growth. In fact, the central city and competing city B begin to merge very early in the simulation, at t = 65 (Figure 41). A significant number of clusters are settled early in a simulation run, and these incubate a volume of internal growth that diffuses within the system. This is roughly equivalent, in a sense, to similar phenomena in real-world contexts, e.g., in examples such as the 'Silicon Valley' experience in Palo Alto and the Bay Area of Northern California, and similar patterns in the Seattle-Tacoma area of Washington, as well as Salt Lake City and Ogden in Utah. In each of these cases, relatively sleepy peripheral areas gain some form of innovative advantage that establishes a future base for impressive growth—Palo Alto and Santa Clara in the California example, the Wasatch front in the Utah example, and the Redmond and Bremerton areas in the Washington example. Growth in the simulation is still cut off 75% of the way through the simulation run, but at that stage there is more than enough internal momentum in competing cities A and B, and the cut-off has relatively little impact, compared to its use in simulation one.

This is much like events that take place in many sprawling cities. Once peripheral areas garner enough of a foothold they often incorporate as independent townships, with independent control over localland-use and zoning. Invariably, the status quo—low density sprawl—is protected rather than more compact forms of development.

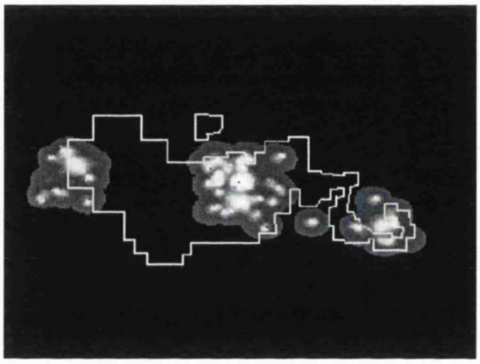
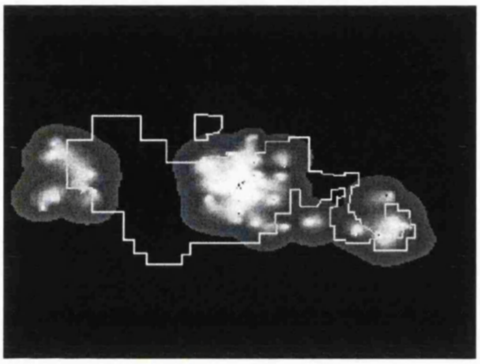
Simulation two generates the same levels of suburbanization as in simulation one, but the generated urban structure is much different. The city-system is surrounded by a buffer of low-density sprawl, as before, but the main urban mass exhibits a much more polycentric structure with *many* well-established cores (Figure 40). This generates a different urban future to that observed in simulation one. In simulation one, low-density peripheral sprawl dominated and it was mentioned that this was synonymous with situations whereby peripheral areas might organize locally—in a politically fragmented manner, as discussed in chapter 6—and reinforce a regime of low-density sprawl. By comparison, growth in simulation two is focused, early on, in peripheral *cores*. While sprawl is marked, it is much more cohesive due to polycentricity in dense core distribution.

The implication for sprawl "costs" would, likely, be significant. The urban pattern in simulation two could be associated with greater system-wide accessibility and potentially lower VMT and vehicle emissions. Simulation one generated a city in which the population living in dense urban settings was roughly equal to that housed in low-density sprawl. If we assume that sprawl dwellers may follow a particular socioeconomic profile commensurate with "white flight" scenarios mentioned in chapter 6, the social justice implications are significant. Simulation one is indicative of large-scale system-wide socio-spatial segregation; simulation two accommodates a potentially more balanced distribution.

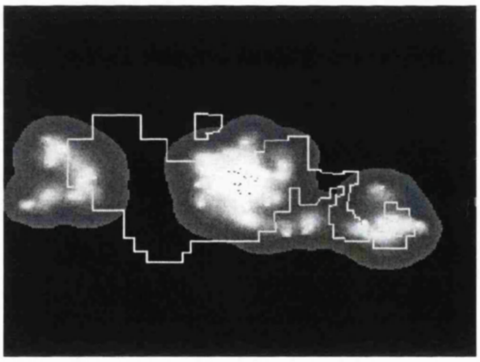
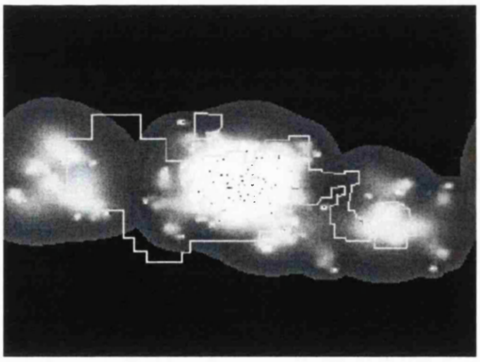
Transition t = 3

Transition t =25

Transition t = 50

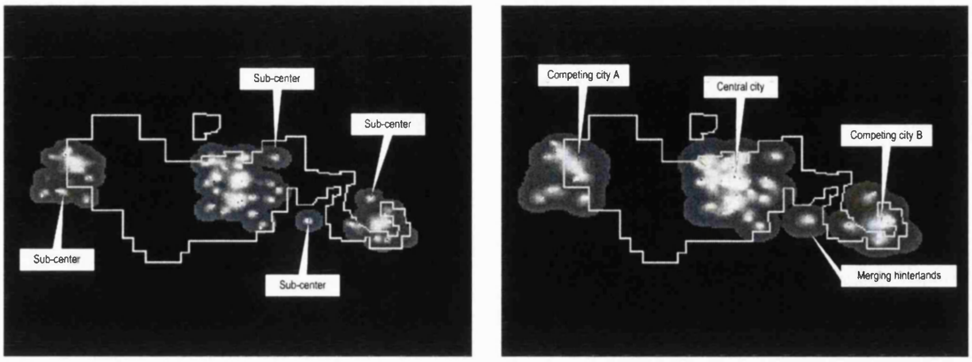
Transition t = 100

Transition t =200

Transition t = 350

Figure 40. The evolution of pedagogic simulation two



(a) (b)

Figure 41. Noticeable features in pedagogic simulation two

(a) Sub-center formation (t = 40). (b) Merging cities (t = 65).

Simulating sprawl in the Midwestern Megalopolis

In the next example, the model outlined at the start of this chapter is applied in a more 'real-world' context to the Midwestern Megalopolis region (Gottmann 1967a) around Lake Michigan in the United States. The area was picked for a number of reasons. It is the area in the United States in which I have lived the longest and is thus quite familiar. Furthermore, Chicago and its surrounding urbanized region, is *the* classic case study city for urban geography, popularized by the 'Chicago School' of the 1930s (Carter 1981). The area provided some unique characteristics for applying the model; in particular, the boundary formed by Lake Michigan.

The model world was coded into discrete GA using a Landsat TM image (Figure 42). Each pixel in the image was coded as an individual cell in a regular lattice structure. The simulated region occupies a 52,125 km^2 area in the real world. The GA lattice comprises a grid, 520 units wide and 630 long—327,600 fixed GA in total, with a real-world resolution of 180,093 m^2 per GA. The Midwestern simulation is specified in much the same way as the abstract simulations described earlier. The simulation is based on the model engine described at the start of this chapter. It is specified as a GAS, with fixed and non-fixed GA. GA are characterized with state variables as before (Table 6). The same general state transition rules, geo-referencing conventions, and movement rules are used. Neighborhood configurations remain the same as in the other simulations. The Chicago simulation is distinct from abstract simulations one and two in its constraint parameters, however.

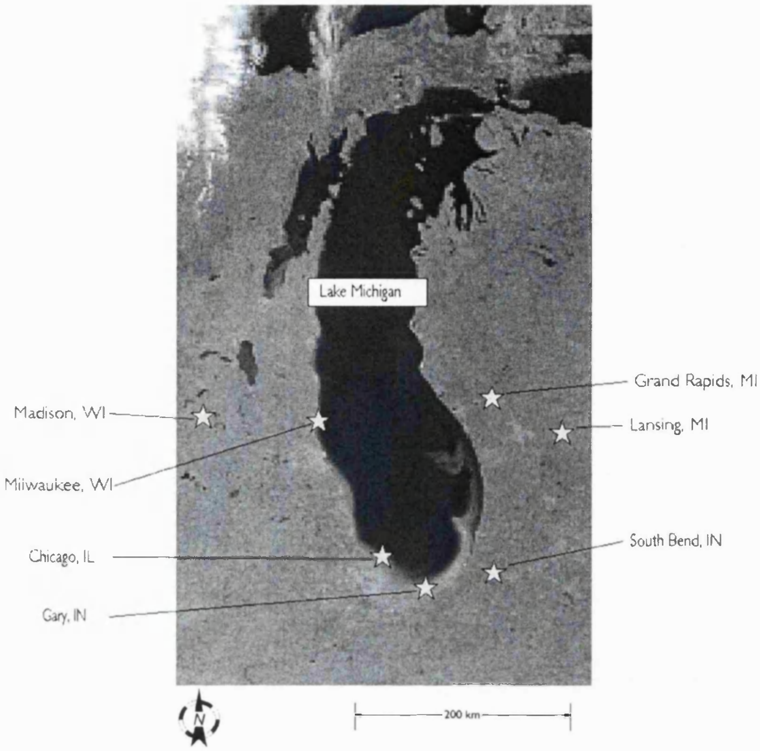


Figure 42. Seed sites in the Midwestern Megalopolis model.

The Chicago simulation is constrained geographically through the introduction of known seed sites for development. The seed sites are specified with respect to those locations in the area that came to dominate as urban centers in the region—namely, the city centers with the largest current population (Table 7, Table 8). Seven such sites were identified and introduced: Madison, WI; Milwaukee, WI; Chicago, IL (Figure 43); Gary, IN; South Bend, IN; Lansing, MI; and Grand Rapids, MI (Figure 42). Each of these sites served as a gateway for the introduction of population to the simulation, thereby ensuring that the simulation retained some basic regional (and geographic) similarities with conditions in the real world.

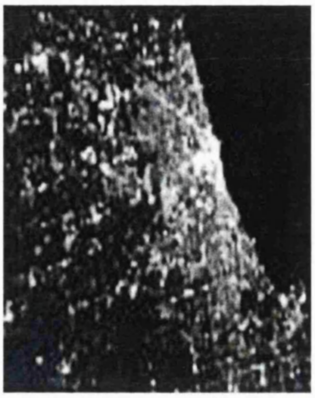
Table 7. Population for the simulated cities, 1980 to 2010[6]

City	1980	1990	1996	Proj. 2000	Proj. 2005	Proj. 2010
MAD	323,545	367,085	394,487	419,800	442,500	464,700
MIL	1,397,020	1,432,149	1,453,050	1,497,700	1,535,600	1,576,300
CHI	7,246,048	7,410,858	7,726,089	8,029,600	8,328,400	8,626,300
GRY	642,733	604,526	621,132	624,500	629,600	639,200
SBD	241,617	247,052	257,338	266,600	275,400	284,900
GDR	840,824	937,891	1,016,273	1,041,000	1,088,000	1,134,500
LAN	419,750	432,684	446,820	459,300	474,100	489,900
ALL	11,111,537	11,432,245	11,915,189	12,338,500	12,773,600	13,215,800

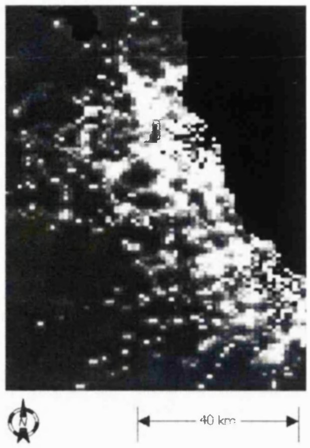
Table 8. Population density and population change, 1980 to 1997 (Source: U.S. Bureau of the Census.)

City	Density 1997 sq. mi.	% Δ 1980 to 1990	% Δ 1990 to 1997
MAD	330.7	13.5	8.3
MIL	994	2.5	1.3
CHI	1534.8	2.3	4.9
GRY	681.2	-5.9	3.1
SBD	564.3	2.2	4.5
GDR	372	11.5	9.4
LAN	262	3.1	3.4
ALL	677 (average)		

6 (Proj. refers to projected population. MAD: Madison, WI MSA; MIL: Milwaukee-Waukesha, WI PMSA; CHI: Chicago, IL PMSA; GRY: Gary, IN PMSA; SBD: South Bend, IN MSA; GDR: Grand Rapids-Muskegon-Holland, MI MSA; LAN: Lansing-East Lansing, MI MSA; ALL: all cities.) (Source: U.S. Bureau of the Census.)



(a) Observed Chicago



(b) Simulated Chicago

Figure 43. Observed and simulated conditions in Chicago[7]

The simulation is constrained in one additional way, and this relates to both geography and rates of change in the model. As in the pedagogic simulations, the growth rate was specified by means of the introduction of non-fixed—settler-developer—GA to the simulation via fixed gateway GA. In the Midwestern example, growth enters through the city center sites described before. The *volume* of growth introduced at each time step is designed to roughly match known growth values for the particular cities (Table 7, Table 8). The growth rates were varied for different simulation runs to examine the patterns generated, but in the run illustrated in Figure 44, growth rates were scaled relative to known growth in the real world cities. Developer-settler GA originating from these gateways are geo-referenced to the sites through which they are introduced. A greater volume of growth was introduced through Chicago, relative to the other cities; Milwaukee had more growth than Madison, etc. This ensures that the rate of evolution in the simulation is plausible and allows the simulation exercise to focus on the relative impact of the general state transition rules and movement rules in the model.

[7] (a) The pattern of urbanization as revealed by night lights (source: NASA, http://science.nasa.gov/headlines/images/lights/chicago_lights.jpg); (b) a section of the simulated world corresponding to the Chicagoland area

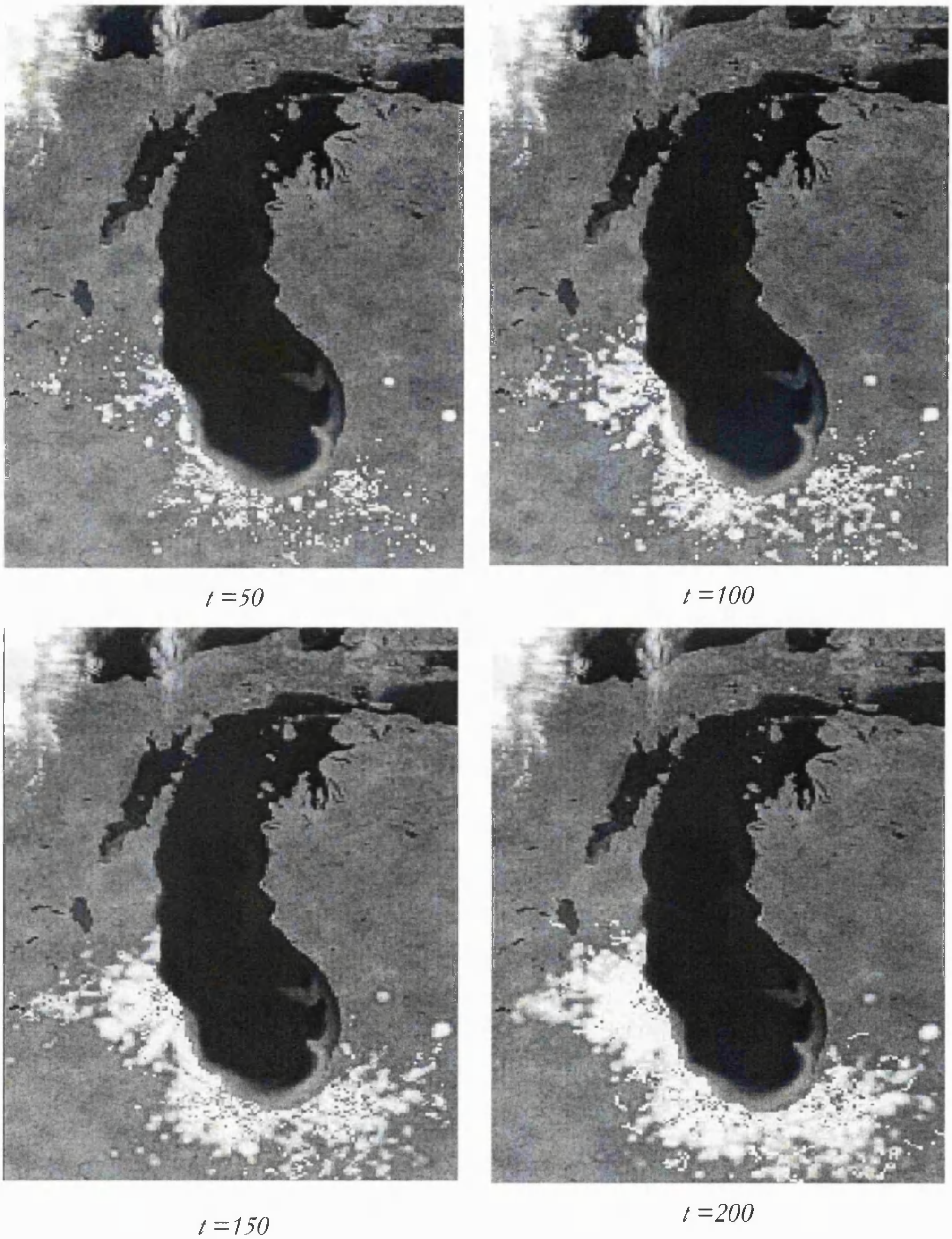
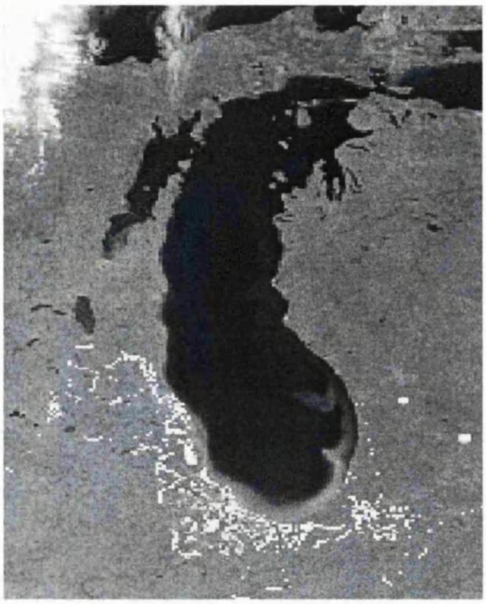


Figure 44. Simulated Midwestern growth at various stages

(a) Decentralization without end



(b) Growth under the road-like and irregular rules

(c) Growth under the leap-frog rule

(d) Growth under the immediate, nearby, and leap-frog rules

Figure 45. Simulated Midwestern urbanization under different scenarios

Using these specifications, the simulation was run with varying parameters. The example illustrated in Figure 44 was run with equal weighting of transition rules, for two hundred iterations, from a state of only minor settlement in the seed sites (roughly synonymous with conditions in the area at the turn of the Nineteenth Century). These specifications generated a realistic urbanized area. (A more empirical validation exercise was undertaken; the results are discussed in later sections.) The simulated city-system began developing as a loose constellation of urban clusters, scattered in the immediate vicinity of the seed sites identified in Figure 42. By t = 50, the relative dominance of Chicago and urbanized lower Wisconsin is evident in the system (Figure 44). By t = 100, the city-system has begun to coalesce, with road-influenced fingers of growth connecting spatially separated spheres of development. By t = 150, the system has begun to sprawl—with fragmented and lower-density settlement—on the urban periphery, while expanding into previously undeveloped areas. By t = 200, low density areas occupy as much area in the model world as sites with relatively higher densities.

In another simulation, the model was run far ahead into the future as a speculative exercise to examine what the pattern of urbanization might look like if growth continued unchecked (Figure 45a). The end-result was decentralization without end, reminiscent of situations written about by Prof. Sir Peter Hall in the 1980s (Hall 1983).

The relative impact of movement (development) rules was also tested. Simulations were run in which the weighting favoring the enactment of movement rules were adjusted relative to other rules. Used together with equal weights, the rules generate plausible urban patterns, at least visually comparable with those observed in the real world (Figure 43). (Sutton (1997) has demonstrated that night lights are a good proxy for development density in American cities.) This suggests that the rules are capable of generating reasonable urbanization scenarios in the GAS. By emphasizing one or more movement rules over others, it is possible to explore potential growth scenarios under alternative development regimes.

Setting the road-like and irregular rules as the prevailing force in a simulation generates a pattern dominated by linear strips of urbanization (Figure 45b). Density within those strips is relatively high, but the overall pattern of growth is very scattered, with infill only occurring in areas where there is a dense network of strips in physical proximity to each other.

Emphasizing the leap-frog rule relative to other rules generates an altogether different pattern of urbanization, dominated by small isolated clusters of dense settlement, with little to bind them within the urban system (Figure 45c).

Combinations of clustering rules—the immediate, nearby, and leap-frog rules—leads to a very polycentric urban structure, characterized by a tight jigsaw of urban clusters, loosely merged by their respective bands of peripheral low-density hinterland (Figure 45d).

Validating the models

The models were validated by analyzing the patterns produced by the simulations: visually, as discussed, but also empirically. Specifically, landscape metrics were calculated for the simulated cities.

Landscape metrics rely on *patches* and measure various configuration and composition characteristics of patches and landscapes. Patches represent discrete (spatial) areas or (temporal) periods of relatively homogeneous conditions. The concept can be applied to urban landscapes in the context of representing discrete areas of land cover or land-use (Alberti 2001; Alberti & Botsford 2000). Patches of different classes of variable (e.g., state values) can be judged relative to other similar or diverse patches within a larger landscape composed of mosaics of such patches. Landscape metrics are just beginning to be used as a validation scheme for urban automata models (see Herold 2002; Torrens 2004b for examples).

Landscape metrics are used to measure two major characteristics of landscapes—composition and spatial configuration (Turner 1989). Composition refers to the presence and amount of different patch types within a landscape, without explicit reference to their spatial features. Configuration refers to the spatial distribution of patches within a landscape.

Configuration metrics have advantages as a measure of sprawl, providing an index of the amount of space-filling and fragmentation in a city's urban pattern (Torrens &

Alberti 2000). Three configuration measures are used here, to assess the degree of sprawl in simulated scenes, each at a landscape scale.

Perimeter-Area Fractal Dimension (PAFRAC) measures the extent to which patches fill a landscape. High values of PAFRAC denote situations in which patches fill-up a space; low values are synonymous with cases in which patches fill space to a lesser extent. A PAFRAC value greater than one for a two-dimensional landscape denotes a departure from Euclidean geometry and an increase in patch shape complexity. Differences in PAFRAC value can suggest differences in the underlying pattern-generating process (Krummel *et al.* 1981). PAFRAC approaches a value of two for landscapes with highly convoluted geometries. PAFRAC ranges in value from one to two, and is calculated using the slope of a regression line obtained by regressing the log of patch area against the log of patch perimeter. It is calculated as a double-log fractal dimension,

$$PAFRAC = \frac{2}{\dfrac{\left[N\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\ln p_{ij} - \ln a_{ij}\right)\right] - \left(\sum_{i=1}^{m}\sum_{j=1}^{n}\ln p_{ij}\right)\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\ln a_{ij}\right)}{\left(N\sum_{i=1}^{m}\sum_{j=1}^{n}\ln p_{ij}^{2}\right) - \left(\sum_{i=1}^{m}\sum_{j=1}^{n}\ln p_{ij}\right)^{2}}} \qquad \text{Eq xl}$$

In the formula above, a_{ij} is the area of patch *ij*; p_{ij} is the perimeter of patch i, and N is the total number of patches in the landscape.

In urban contexts, fractal dimension is illustrative of the space-filling ability if a city; the higher the dimension, the less scattered a landscape might be considered.

Contagion is the probability that two randomly-chosen adjacent cells belong to the same class (state). It is calculated on a cell-by-cell basis, rather than a patch-by-patch basis. Contagion is the product of two probabilities: the probability that a randomly chosen cell belongs to category type *i*, and the conditional probability that, given a cell belongs to category *i*, one of its neighboring cells belongs to category *j* (McGarigal & Marks 1995). Where contagion is low, a landscape can be said to be comprised of many small and dispersed patches, i.e., fragmented. High contagion values are indicative of more 'compact' landscapes. Mathematically, contagion is calculated using the following formula (Li & Reynolds 1994),

$$C = \left[1 + \frac{\sum_{i=1}^{m}\sum_{j=1}^{m}\left[\left(P_i \frac{g_{ij}}{\sum_{j=1}^{m} g_{ij}}\right) * \left(\ln P_i \frac{g_{ij}}{\sum_{j=1}^{m} g_{ij}}\right)\right]}{2\ln(m)} \right] * 100 \qquad \text{Eq xli,}$$

C is the contagion, P_i is the proportional abundance of category type *i*, g_{ij} is the number of adjacencies between cells of category type *i* and all other category types, and m is the total number of category types.

The *Interspersion and Juxtaposition Index* (*IJI*) measures adjacency on a patch-by-patch basis. Higher values are synonymous with landscapes in which patch types are well interspersed (equally adjacent to each other). Lower values occur when landscapes contain patches that are poorly interspersed (there is a disproportionate distribution of patch type adjacencies). When *IJI* is zero in value, there is an uneven distribution of adjacencies between patch types. A value of 100 is indicative of a situation in which all patch types are equally adjacent to each other (McGarigal & Marks 1995). High values of *IJI* thus represent a relatively greater degree of homogeneity in a landscape. *IJI* is expressed as a percentage, and can be calculated using the following formula,

$$IJI = \frac{-\sum_{i=1}^{m}\sum_{j-i+1}^{m}\left[\left(\frac{e_{ij}}{E}\right) - \ln\left(\frac{e_{ij}}{E}\right)\right]}{\ln\left(\frac{1}{2}\left[m(m-1)\right]\right)} \cdot 100 \qquad \text{Eq xlii}$$

IJI is the value of the Interspersion and Juxtaposition Index. e_{ij} is the total length of edge in the landscape between patch types *i* and *j*, including landscape boundary segments representing true edge only involving patch type *i*. *E* is the total length of edge in the landscape. m is the number of patch types in the landscape.

Results

The results of the simulations were output at their last iteration, as .png image files. These files were then converted to color .bmp files, using the *imageJ* image processing package (Rasband 2003). The 'mode' and 'adjust' filters in *Adobe Photoshop* were then used to further process the output, converting color .bmp files to monochrome and then inverting the images. Eight-bit binaries were then run through *FRAGSTATS* (McGarigal & Marks 1995; McGarigal *et al.* 2003), with the background of the images (undeveloped sites) excluded from the analysis. Total patch numbers, PAFRAC, Contagion, and IJI were calculated at a landscape level, using an eight-pixel (Moore) neighborhood for judging adjacency. The raw ASCII files generated in *FRAGSTATS* were then parsed, and landscape metric values were extracted.

Abstract simulations

Figure 46 to Figure 52 illustrate the results of the validation exercise. The two abstract simulations demonstrate very different sprawl-like characteristics. Simulation one generated more patches—spatially distinct 'blobs'—than the polycentric simulation two. The patch total was 14,375 for simulation one, and 3,066 for simulation two (Figure 46). This indicates that simulation two was relatively less fragmented than its counterpart. The values for PAFRAC support this contention. Simulation one had a fractal dimension of 1.5305; the value for simulation two was higher at 1.5321 (Figure 47). Both of these values are commensurate with the fractal dimension of cities in real-world contexts (See Herold & Clarke 2002 for values for Santa Barbara in California). The higher value for simulation two also suggests that the simulated city in that experiment did a better job of filling the space it occupied, although the values are not dramatically different. The values for contagion further support the hypothesis that the two simulations generated cities with different spatial structures and patterns of sprawl. There is a dramatic difference in the percentage of contagion recorded for the two simulations (Figure 48). Simulation one yielded a contagion value of 48%, while the figure for simulation two was much higher at 65%. Higher contagion is indicative of a greater degree of compaction of cells in a landscape. The city generated in simulation two can thus be considered less sprawling than its counterpart in simulation one. The results for Interspersion and Juxtaposition produced similar results (Figure 49): simulation one demonstrated a relatively high IJI (54%), indicative of a landscape in which patches are well-interspersed. Simulation two produced a much lower IJI (37%), suggesting poorer interspersion between patches. The higher value of IJI for simulation one suggests that landscape is more homogeneous than that generated by simulation two.

Overall then, the city generated by simulation two—the polycentric simulation—can be regarded as more compact and less sprawled than that generated in simulation one.

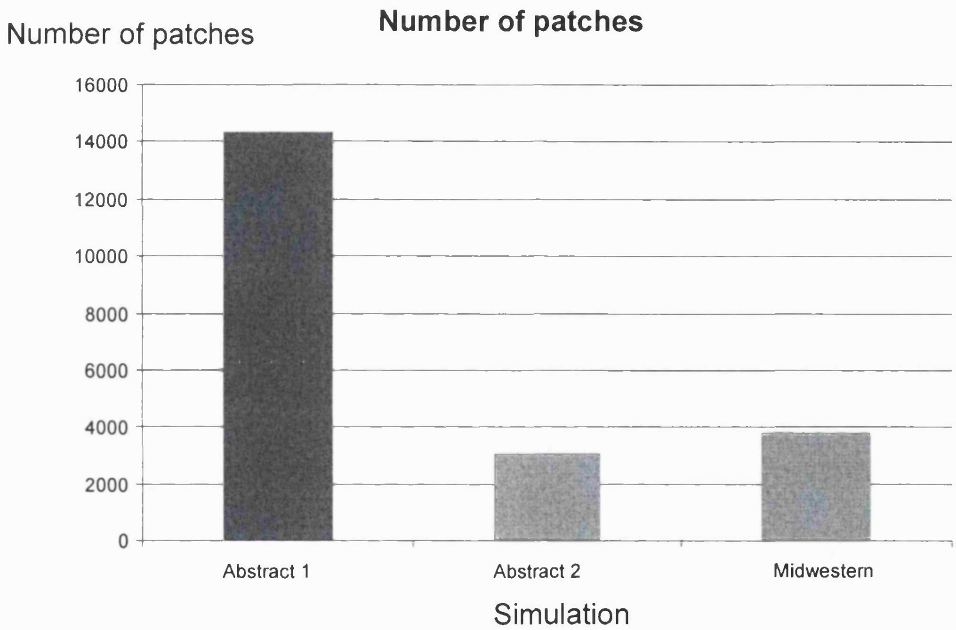


Figure 46. The number of patches at the end of various model runs

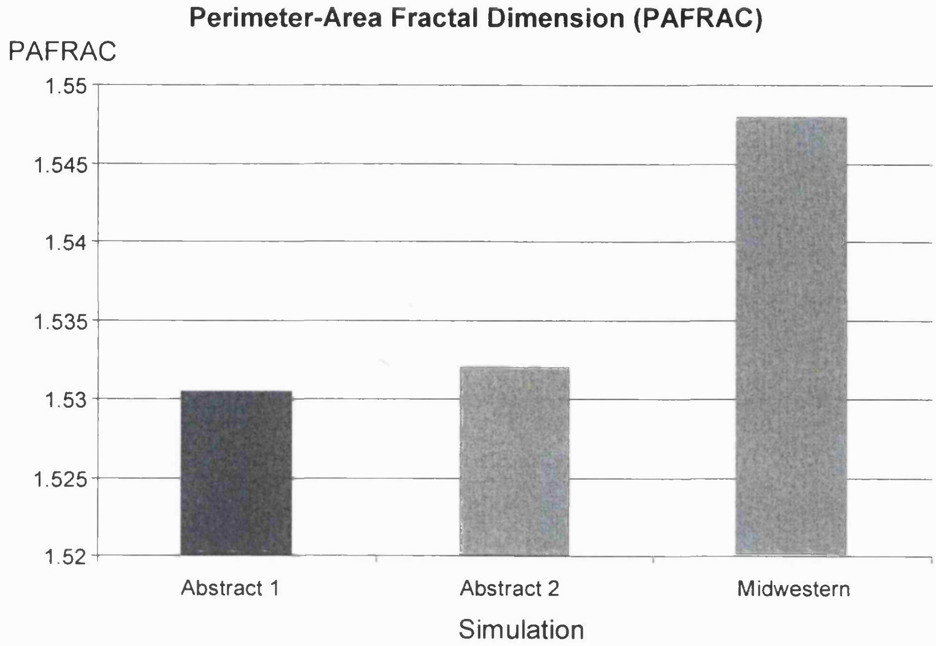


Figure 47. Perimeter-area fractal dimension at the end of various model runs

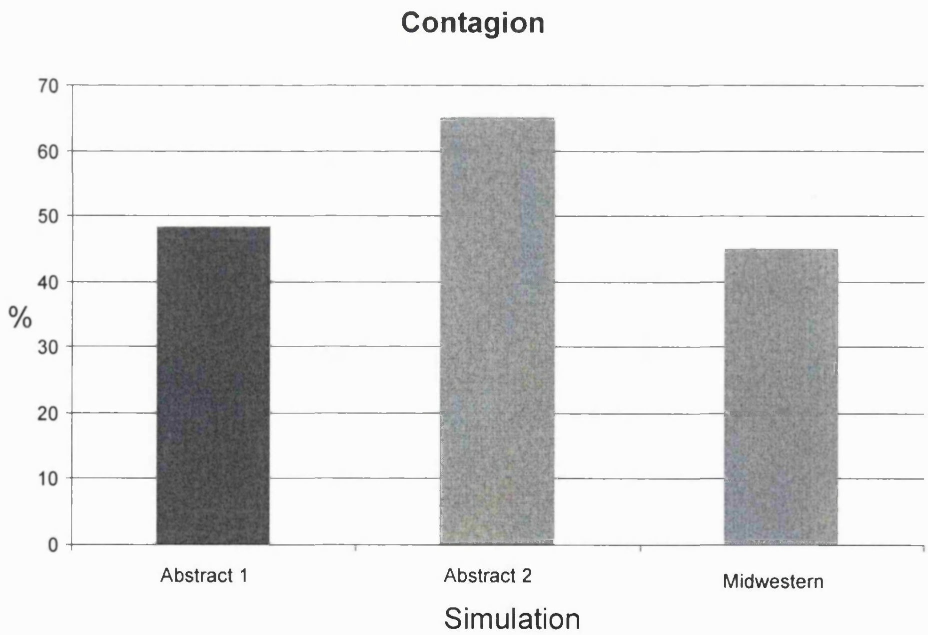


Figure 48. Contagion at the end of various model runs

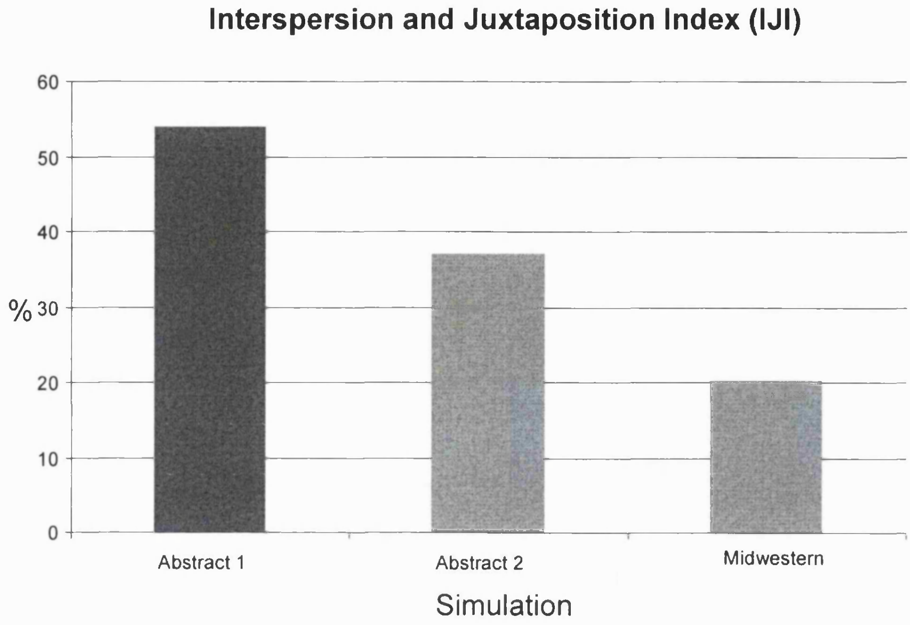


Figure 49. Interspersion and juxtaposition at the end of various model runs

Midwestern simulation

The results for analysis of the Midwestern simulation were consistent with expectations. The simulation generated a realistic urban landscape. This suggests that the model described at the start of this chapter is robust to changes in parameterization and seed conditions.

The number of patches generated by the Midwestern model was consistent with abstract simulation two—the model produced 3,782 patches by the end of the simulation run (Figure 46). The fractal dimension was also consistent with the pedagogic simulations, and with real-world cities, at a value of 1.5479 at the end of the run (Figure 47). The degree of contagion was 45% (Figure 48), while the amount of interspersion and juxtaposition was reported as 20.15% (Figure 49). The contagion score was low relative to the pedagogic simulations, suggesting that the Midwestern model generated a more sprawl-like landscape. Interspersion and juxtaposition was much lower than in the pedagogic simulations. This is indicative of relatively lower homogeneity in the landscape, again an indicator of sprawl.

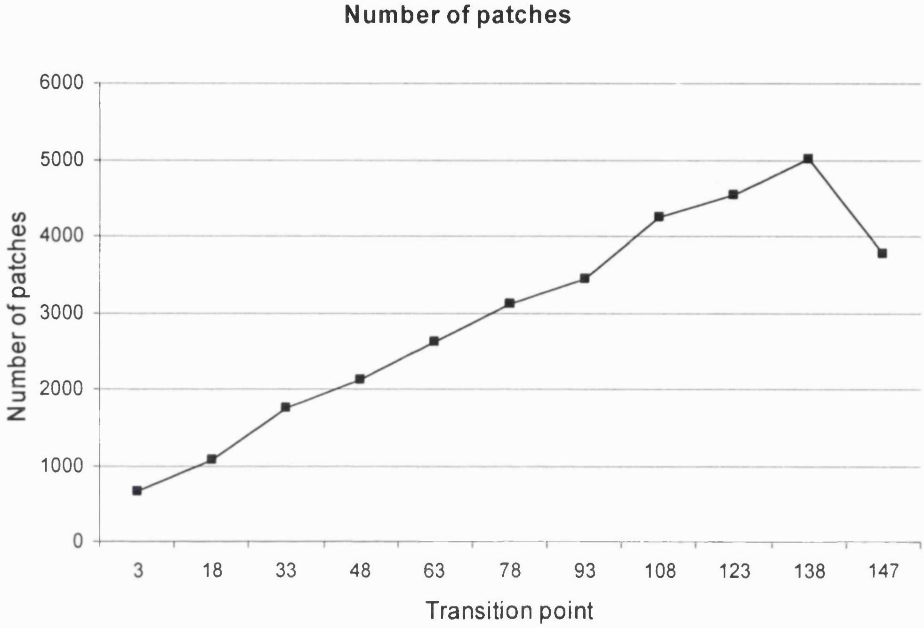


Figure 50. Change in the number of patches in the Midwestern simulation

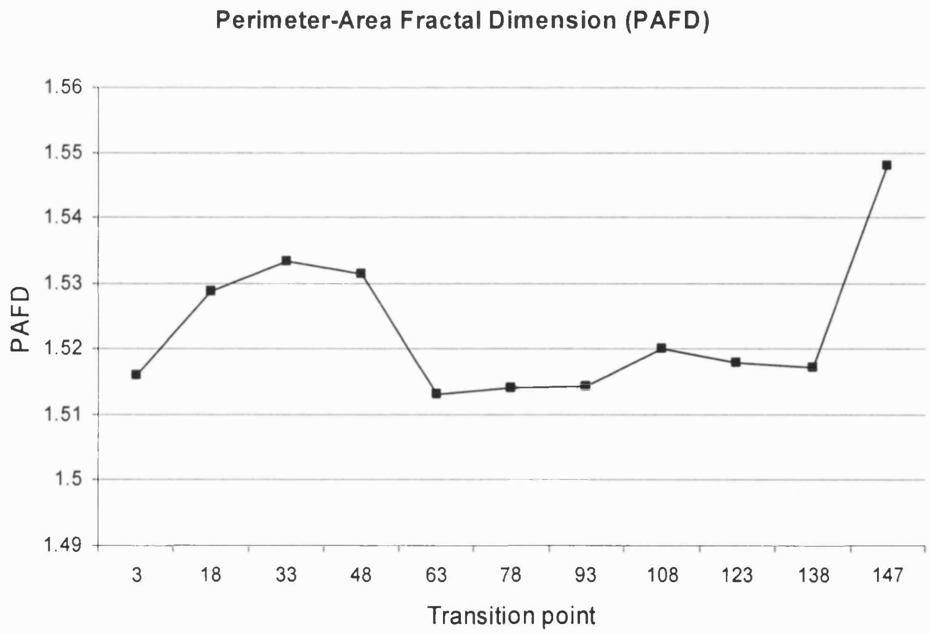


Figure 51. Perimeter-area fractal dimension change in the Midwestern simulation

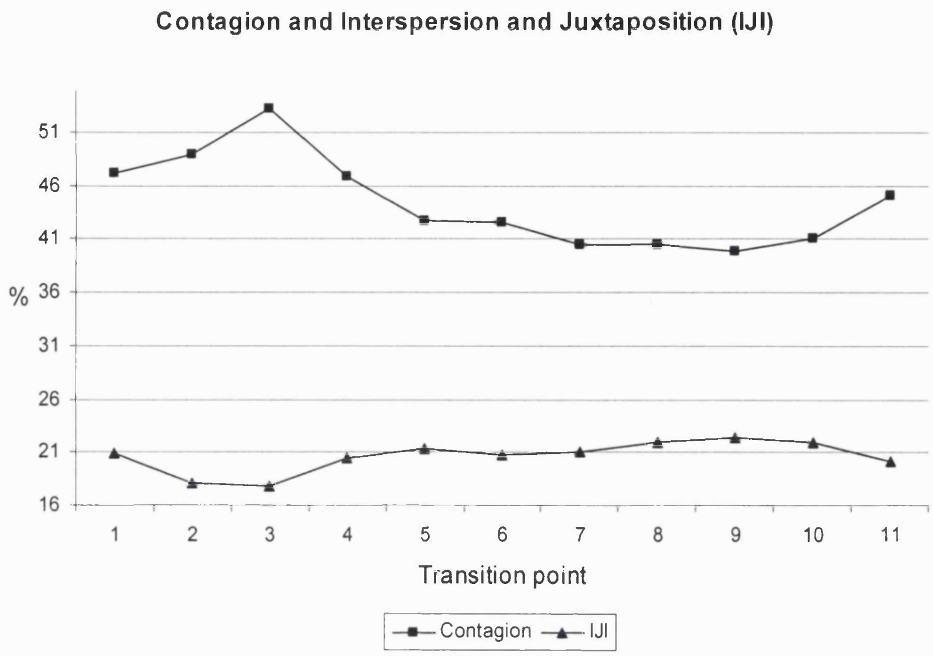


Figure 52. Contagion and interspersion change in the Midwestern simulation

The analysis was performed across the life-time of the simulation run for the Midwestern simulation, to explore changes in the structure of the simulated city as it evolved within the simulation. Analyzed across time-steps, the results indicate a significant cyclical development to the simulated city, with rapid changes in initial conditions, followed by a period of relative stability, and a sharp transition toward sprawl at the end of the model run. This is consistent with the 'lifecycle' stages of an urban system referred to before, whereby cities go through periods of relative compaction, expansion, and decentralization (Hall 1983).

The number of patches showed a progressive increase, from just a handful of seed sites to over 5,000 by t = 138. After that point, the number of patches started to decline steadily, as fragmented areas began to coalesce (Figure 50). The fractal dimension fluctuated over the course of the simulation run, although it remained within a reasonable range, as compared to dimensions for other cities that have been mentioned in the literature (Batty & Longley 1994). This suggests that the simulated city went through cycles of growth, with rapid space-filling at the beginning of its evolution, followed by a period of relatively stable growth. The value climbed markedly toward the end of the simulation run as the simulated city began to sprawl at a growing rate (Figure 51). Contagion and interspersion and juxtaposition demonstrated an almost inverse relationship over the simulation run. The degree of contagion in the landscape grew early in the simulation, declining thereafter before climbing rapidly toward the end of the model run (Figure 52). This is consistent with the results suggested by the other metrics—the city went through an early growth phase dominated by compaction. The decline in contagion thereafter is indicative of relative sprawl. The value of interspersion and juxtaposition in the simulation started off quite high, and subsequently declined quite rapidly, before rising in value, mostly, throughout much of the simulation run (Figure 52). Once again, there was a sharp change at the end of the model run, where the value dipped to its lowest level. This suggests that the simulated city started off with relatively homogenous conditions, losing homogeneity thereafter and entering into a sustained period in which there was poor interspersion. Toward the end of the simulation, there is a strong tendency for interspersion, with a growth in homogeneity, which we can associate with sprawl.

Implications for understanding sprawl

The results of the simulation exercises have a number of implications for understanding sprawl. In terms of the causes of sprawl, each of the factors discussed in Chapter 6 appear to be important in driving sprawl. Moreover, their combined influence is particularly significant.

What is evident from each of the simulations discussed in this chapter is that sprawl is, to a certain extent, inevitable. It is the likely end-state in the natural evolution of a city-system. This is obvious in the context of most well-established cities in the United States. However, it is particularly important in the context of newly-forming cities, such as those developing and growing rapidly in previously-termed "Sun Belt" cities, predominantly situated in the Southwestern region of the United States. For these cities, there is a propensity for urban evolution to jump or skip the natural evolution process, fuelled by higher-than-average growth rates (Table 4) and contemporary development regimes, and go straight to sprawl. However, there is also opportunity to plan cities in such a way that this situation does not occur. The simulations described in this chapter suggest a few—geographic—ways in which policies could be developed to mitigate circumstances.

Sprawl is inevitable if growth occurs. This is evident in the real-world; the fastest-growing cities are generally associated with the highest degree of sprawl (OTA 1995). It is also apparent in the simulations described in this chapter. The message, then, is this: If you want to stop sprawl, stop growth. (Of course, no urban manager wants to hear this, certainly not mayors!) We did just that in abstract simulations one and two; growth was cut-off to the auxiliary cities in the urban system roughly 75% of the way through a simulation run. (This is also the stage in the Midwestern simulation at which the most rapid period of sprawl begins in the cities' evolution.) But, enough decentralization had occurred at the stage that there was sufficient momentum for sprawl anyway. Natural, local-scale diffusion continued the expansion of low density urbanization into the urban periphery in the absence of growth. This is also the case in the real world. In so-called "Rust Belt" cities of the United States, where population (and economic) growth are declining—Buffalo, NY is an example—sprawl continues on the urban periphery. The implication, then, is that sprawl is even more inevitable than we might suspect. This is certainly true in the simulations described here, given that the geographic mechanisms in the model operate as they are designed to do, i.e., given that leap-frogging, decentralization, and local-scale diffusion take place.

This suggests some ways in which sprawl might be tackled, *geographically*. Unchecked growth leads to low-density, blanket sprawl (Figure 45a), with all the associated costs discussed in Chapter 6 implied. Essentially, controlling sprawl requires mechanisms to *manage* growth, sustainably. A number of mechanisms are understood to drive sprawl, and several of these are represented in the simulations described here. The model can be used as an artificial laboratory to explore 'what-if' scenarios relating to changing the mechanisms suggested to cause sprawl. We are most interested in geographic scenarios, and the results of the simulation exercises advocate some options.

Encouraging polycentric development appears to be one solution—allowing leap-frogging, but encouraging sustainable and compact independent clusters, in close proximity (Figure 40, Figure 45d), rather than isolated patches (Figure 45c). Edge cities (Garreau 1992) may be one way to achieve this; Transit Villages (Cervero 1998) are a more likely—sustainable—option. It is important to actually *permit* sprawl to occur, *locally*, on the periphery of these clusters, to facilitate infill and to avoid by-passing large areas of land. This idea is reminiscent of much older theories of urban development, notably, the idea of central place theory (Carter 1981).

Road-like growth can also be used effectively, to link isolated clusters (Figure 39d). (Transit-oriented development may have even greater potential.) However, care must be taken to avoid isolated linear development—ribbon sprawl (Figure 45b).

Conclusions

This chapter has demonstrated the application of the general GAS methodology to the simulation of sprawl. The framework is particularly beneficial in modeling sprawl, allowing for the description of system dynamics as a function of spatial interactions between mobile, agent-like entities and a static, CA-like environment. Moreover, the framework allowed for the generation of very realistic macro-scale urban structures from these local-scale mechanisms.

The simulations described in this chapter were developed as artificial laboratories for exploring the relative—and geographic—impact of the proposed causes of sprawl outlined in earlier chapters. The model generated sprawl-like cities in each of the simulation runs, and by varying the influence of rules within the model, facilitated exploration of the potential drivers of sprawl.

After validating the model through the use of fractal analysis and metrics from landscape ecology, various potential options for managing sprawl were inferred. The results suggest that sprawl might best be tackled—geographically—by encouraging compact and sustainable clusters of leap-frog development in close proximity; sprawl on the periphery of these clusters then serves as an in-fill mechanism, rather than continuing on the periphery of a larger urban mass in an unsustainable fashion. Moreover, it was determined that road-influenced growth could help to link isolated fragments of sprawl on the urban periphery under certain conditions.

The simulations discussed in this chapter were designed to explore geographic dimensions of sprawl, focusing on mimicking the spatial distribution of growth in dynamic contexts. From the discussion in Chapter 6, however, it is clear that there are other important components to the sprawl phenomenon that these simulations have not addressed—namely, preference-based drivers. The next chapter describes another modeling exercise, once again based on a GAS framework that facilitates the evaluation of preference-based behavior in an artificial residential submarket, roughly equivalent to a single fixed infrastructure GA in the growth-based simulations of this chapter. In this example, non-fixed GA are used to simulate relocating households, while fixed GA represent real estate in a hypothetical neighborhood. The rules driving dynamics in the model focus on the evaluation of household preferences in the simulated environment.

Chapter 9. Residential mobility within-the-cell

"He closed his eyes, not wanting to see the new buildings. But they were still there, in the darkness and the light behind his lids. And as he watched, they slid apart, deliquesced, and trickled away, down into the mazes of an older city." (Gibson 1996, p.83)

Introduction

The model described in the last chapter was designed to simulate geographic components important to sprawl dynamics, from a growth-distribution perspective and at a synoptic level across a regional city-system. The model introduced in this chapter takes a different approach. The model has been designed with consideration of the premise that preferences have a role to play in urban dynamics in phenomena such as sprawl; specifically, that sub-markets in sprawling areas transition from the bottom-up through the interaction of households in a residential real estate market. Essentially, the model described in this chapter focuses on the dynamics within a cell from the growth model of sprawl described in the last chapter. The model was built as a hypothetical extension to the growth model described in Chapter 8, although the two are not linked. A simulation begins with the assumption that a quota of "settlement" has been delivered from the previous growth simulation to a residential submarket in this model. Once again, the model is designed around the Geographic Automata Systems (GAS) framework outlined in Chapter 5.

The modeling exercise described in this chapter was designed with two main aims. First, as a modeling exercise, it demonstrates the application of the GAS framework to more local-scale geographies: individual households, individual properties, communities of residents, and real estate submarkets. Also, it demonstrates the ability of the GAS framework to support cross-scale spatial dynamics—communities and submarkets are treated as systems in their own right, with the ability to change and evolve dynamically as the entities within them interact. In a sense, this demonstrates the ability of GAS to support cross-scale emergence, although emergence is a notoriously murky concept in modeling terms (Epstein 1999; Faith 1998; Horgan 1995).

Second, the model is designed to shed light on the mechanisms driving sprawl at *very local-scale* geographies. As the discussion in part two of the thesis illustrates, much of the explanatory work investigating the causes of sprawl has focused on coarse spatial resolutions—the region and the city; at best, at the intra-urban level. The distribution model described in the last chapter simulated many of those processes (scatter, leap-frogging, transport-influenced growth, polynucleation, etc.). Many of the preferences that drive sprawl occur at the intra-urban level; households choose to live centrally or in the suburbs. Much of this is implied in the growth model discussed in the last chapter, with characterization of developer-settler geographic automata. However, community-level and very local-scale dynamics are also significant. In particular, community resistance or resilience to change has long been understood to influence the socioeconomic—and sometimes ethnic—characteristics of suburban areas. This is covered in the literature, with reference to 'white flight', socio-spatial inequalities, and exclusionary practices (see Knox 1989; Palm 1982 for reviews in American contexts). In terms of sprawl, the focus of this model is on determining how suburban 'cells' become resilient to socio-economic change, from the bottom-up.

Put succinctly, the model works in the following way. Households move into and out of submarkets and individual properties using realistic behaviors and exercising preferences. As interactions between households and properties evolve dynamically over the course of a simulation run, the nature of the submarket itself changes, from the bottom-up. Households may then react to these changes, altering their behavior from the top-down. The influx of new households to the system is treated exogenously, as a data feed to the simulated system, but triggered by endogenous events within the modeled system. The characteristics and mechanisms used to design the model stem from residential mobility theory; this is discussed in more detail in the next section. Much of this theoretical background comes from studies external to sprawl contexts (and, as discussed in part one of the thesis, it is not commonly used in practice to drive explicitly spatial forecasts); there has been little work on residential mobility in sprawl-specific contexts, and the author is unaware of any modeling work on local-scale residential mobility and sprawl.

This chapter is organized as follows. In the next section, automata models of local residential dynamics are discussed. This is followed by discussion of the theoretical understanding of residential mobility, and this forms the context for assumptions considered in developing the model and running the simulation described in this chapter. Following that section, the specification of the model as a Geographic Automata System is described. A typical simulation run is then illustrated, demonstrating the inter-play of the various components that make up the model. A number of simulations designed as experiments within the model environment are then described. Finally, the chapter concludes with discussions of the implications of the modeling exercise for simulation in a Geographic Automata Systems environment, and for understanding sprawl.

Automata models of residential dynamics

There have been some automata-based models of residential dynamics developed for local scales, however. Schelling's model of social segregation was developed in the 1970s (Schelling 1969, 1978). It is not strictly an urban model, although it has relevance to cities. It is not strictly automata-based either! Yet, in conceptual terms, it is an excellent example of how urban CA operate. Schelling set out to demonstrate the hypothesis that "the interplay of individual choices, where unorganized segregation is concerned, is a complex system with collective results that bear no close relation to individual intent." (Schelling 1969, p.488) He was concerned with phenomena of large-scale spatial segregation of socioeconomic—particularly ethnic—groups in urban America.

> "The demographic map of almost any American metropolitan area suggests that it is easy to find residential areas that are all white or nearly so and areas that are all black or nearly so but hard to find localities in which neither whites nor nonwhites are more than, say, three-quarters of the total." (Schelling 1969, p.488)

Schelling was interested in finding out why ethnic groups did not live in spatially integrated areas of the city. Schelling proposed a simple model to test various hypotheses about the mechanisms driving metropolitan segregation. Conceptually, this model closely resembles a two-dimensional CA. Cells could adopt three color types (states): (in our example these are) blue[8] ('blue people'), white ('white people'), or gray (vacant sites without population). Each cell carried an additional state that denoted the 'contentedness' of its population. Residents of a given cell were content with their location so long as the *majority* of their neighbors (in a Moore neighborhood) were the same color. If the residents were not content they moved to a new (random and available) location in the next time step (i.e., the color of the cell transitioned to a new state).

If a simulation is designed along these principles, with these simple structural parameters, a simple rule set, and a randomly distributed pattern as a seed condition for the model (Figure 53), you find that after several iterations a pattern of distinct segregation will emerge. Blues will cluster into large homogenous groups, with whites similarly arranged in their own independent clusters (Figure 54). Extrapolating these conditions to a real world context, you end up with a divided city in which there is strong spatial segregation. Color-exclusive clusters develop even though people really do not mind living in integrated neighborhoods. This relates, theoretically, to the idea of the 'tipping balance': a threshold, above which, a system shifts to a new phase and may settle in a new and distinct equilibrium. Such phase shifts are among the common characteristics of complex adaptive systems.

Other automata-based residential dynamics models have been built by Itzhak Benenson and colleagues at Tel Aviv University. Their models are designed to explore the impact of differing preferences for housing among different ethnic groups in Tel Aviv, and they operate at micro-scales whereby individual households are simulated in independent properties (Benenson 1998, 1999; Benenson & Omer 2000a; Benenson *et al.* 2002; Portugali *et al.* 1994).

[8] They may appear as dark grey if you are viewing this text in monochrome.

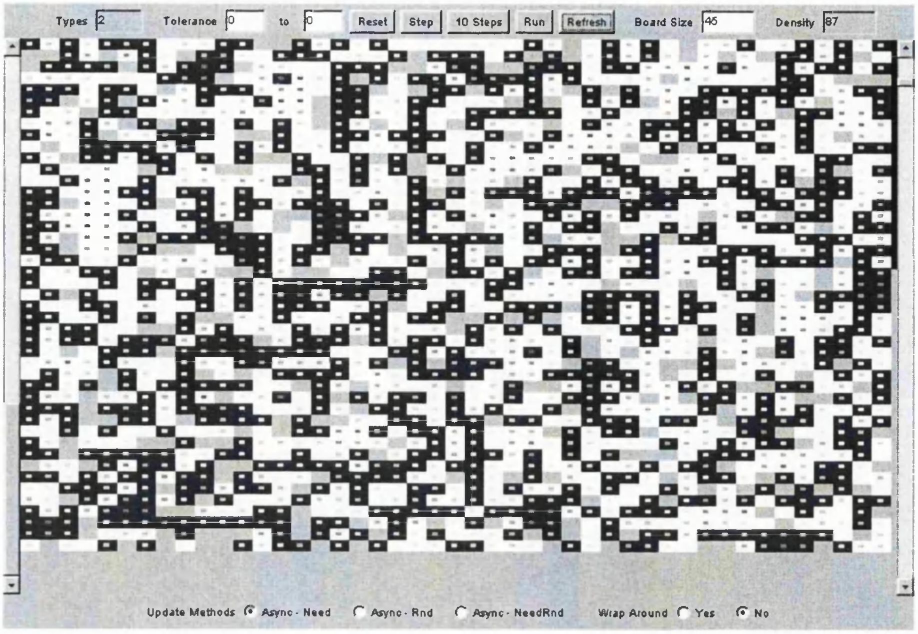


Figure 53. Initial random seed conditions for a Schelling segregation model

(Developed using Yale Wang's Schelling model; source code is at http://www.cs.caltech.edu/~yale/ec126/segregation/).

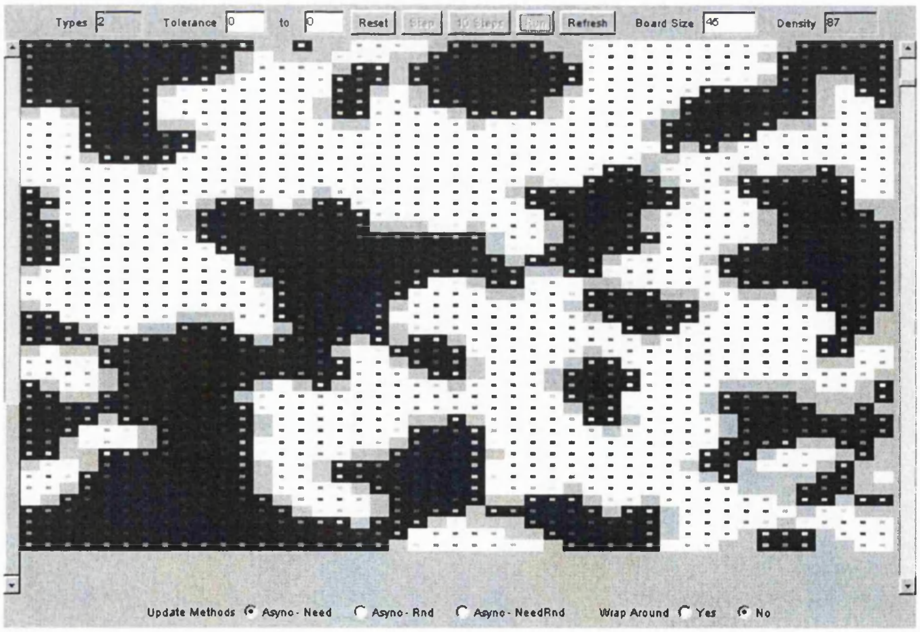


Figure 54. The Schelling segregation model after several iterations

(Developed using: Yale Wang's Schelling model; source code is at http://www.cs.caltech.edu/~yale/ec126/segregation/).

The forces underlying residential mobility

Geographical exploration of the factors leading to residential mobility has been undertaken from a number of different vantage points. A variety of forces—at different scales of observation—can be identified as being important in explaining residential mobility in an urban context (Figure 55). In the model described in the last chapter, attention was focused on the regional scale; in the model described here, attention is focused on very local scales—communities and submarkets. At that scale, some authors have characterized residential search as a hierarchical process, focusing on residential mobility at the scale of the neighborhood, real estate, the household, and the individual (Speare *et al.* 1975). We could add the sorts of intra-urban factors featured in the model in Chapter 8 to the top of that hierarchy.

Traditionally, consideration of intra-urban aspects of residential location choice were focused on bid-rent style associations between household desires to live close to downtown opportunities and their ability to compete with other activities (industrial, commercial) for real estate in that scheme (Alonso 1960; Muth 1969). This bid-rent framework was reviewed in Chapter 2. Similarly, other work has focused on intra-urban residential location in terms of accessibility-related factors, with accessibility to work opportunities, retail and recreation opportunities, and public transit commonly considered (Clark 1970; Handy & Niemeier 1997; Hodge 1986).

There is also a large body of literature concerned with residential location models and methodology as demand-side components to large-scale models of urban systems (see De la Barra 1989; Waddell 2000; Waddell *et al.* 2003; Wegener 1994). For the most part, however, these are economic in nature and do not generally deal with the sorts of local-scale *interactive* dynamics discussed in this chapter. (Although, UrbanSim (Waddell 2000) comes quite close.)

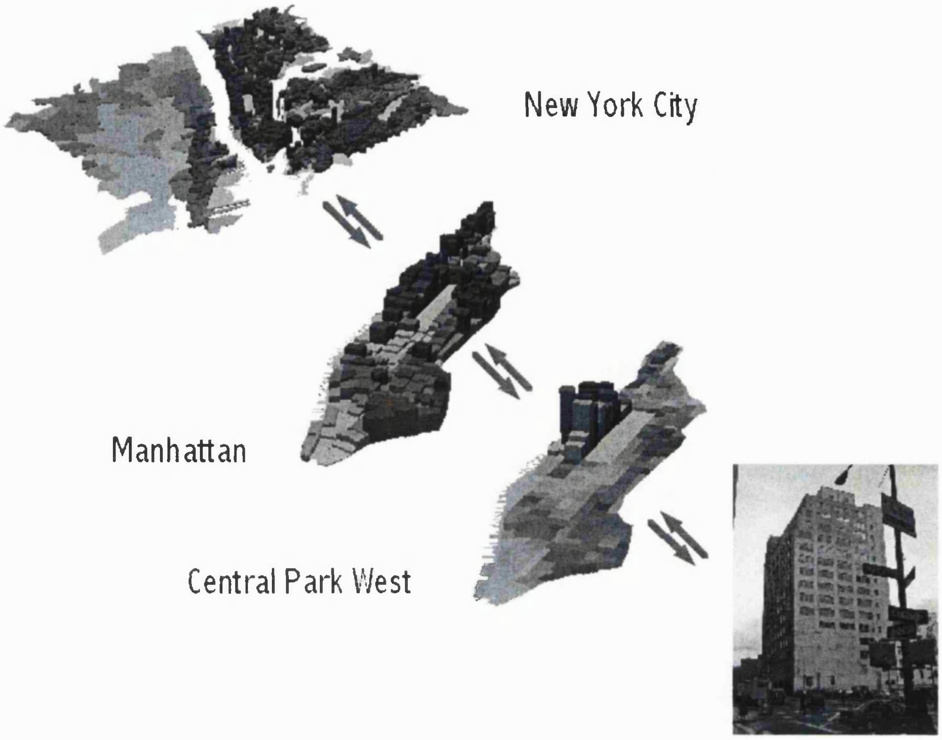


Figure 55. Scaling in the consideration of residential mobility.

At a more meso- level in the urban hierarchy, residential mobility has also been studied from the perspective of area-based search for housing. In the context of the Los Angeles area, Huff (1986) found that households looked for a suitable area in which to target their search, then evaluated properties within that general location. Again, in the context of Los Angeles Clark (1982a) extended that idea, finding that households targeted particular *housing markets or submarkets* as their primary search criterion.

Within housing submarkets, a variety of community-level attributes are considered to be important and we will discuss these later when considering residential preferences. A number of real estate characteristics are also considered to be important, and these are commonly considered—collectively—as 'bundles' of attributes. The relative importance of real estate factors are often determined using hedonic price models, which relate real estate attributes to the market price of the property (see Waddell 2000 for an example).

Invariably, there is feedback between these different scales of observation (Figure 55) (Benenson 1999). For example, deterioration and refurbishment of individual properties in a submarket can lead to ghettoization or gentrification in a spatial submarket (see O'Flaherty 1996 for examples).

Residential relocation is also widely understood to be a two-stage process (Brown & Moore 1970; Clark 1993; Tu & Goldfinch 1996; Wolpert 1965). First, a household decides to move. Second, it relocates to a new home.

Why do people move?

The decision to move is largely governed by stress and resistance factors, a mixture of stimulus and attraction to other housing opportunities.

A great deal of research has been performed, examining the stimuli for residential mobility. The stress-resistance idea is one of the most popular hypotheses. In this framework, the decision to move is understood to be prompted by some 'stress'. Stress then acts as a 'trigger', initiating a mobility decision (Wolpert 1965) or *in situ* adaptation. Clark and Cadwaller (1973) demonstrated that the volume of stress felt by a household can be correlated with the intensity of their desire to relocate. When determining whether to move or not, a household may balance the stress stimulus with some desire to stay or a dearth of better alternatives—'resistance'.

Stressors may take the form of environmental considerations (Wolpert 1966), either from neighboring properties or households or conditions within the submarket as a whole. The nature of stress may also take a variety of forms, from general dissatisfaction (Huff & Clark 1978) to dissonance (Portugali *et al.* 1997). In particular, the role of race and ethnicity biases as environmental stressors has been researched quite thoroughly (for example, see Portugali *et al.* 1994; Schelling 1969; Sermons 2000).

Similarly, different characterizations of resistance have been introduced. Simple inertia is perhaps the most common. For example, Golledge and Stimson (1997) discuss 'cumulative resistancy'—the longer a household remains in a given location, the more their satisfaction with that location, or their resistance, grows. Brown and Moore (1970) distinguish between 'movers and stayers' in this context.

A considerable volume of work relates to the idea of *thresholds* for stress-resistance—the notion that there is some form of 'tipping point', beyond which households may initiate a search for a new residential location. Schelling's models of segregation (Schelling 1969, 1971, 1974, 1978) are one example; the work of Sakoda (1971) is another.

A variety of *internal* stress factors for residential mobility have also been identified. In particular, the passage of households (families, in particular) through a life-cycle—or 'life course'—has been demonstrated as particularly important in governing residential mobility. The idea, here, is that households' preferences for housing and location will vary based on their stage in that cycle. As Rossi (1955, p.9) puts it, findings indicate:

> "the major function of mobility to be the process by which families adjust their housing to the housing needs that are generated by shifts in family composition that accompany life cycle changes" (quoted in Golledge & Stimson 1997, p.467).

Other studies confirm the role played by life cycle events as mobility stressor (Pickvance 1973), with some authors linking certain property types with particular life cycle stages—so-called 'housing careers' (Kendig 1984).

Housing search and relocation

We can identify two aspects of housing relocation events—the process of engaging in a property search, and the choice behavior that governs selection (or not) of particular locations and properties.

Housing search

Housing search is very spatial in nature. The role of accessibility and willingness-to-pay has already been discussed above. Much of the spatial aspects of search disappear at very local-scale geographies; as search focuses on particular properties, there is a tendency to focus on various attributes of real estate and research in this area is largely concerned with the role that household preferences play in choice decisions.

Exploration into the intensity, length, and 'stopping rules' for housing search are other threads in the literature on residential mobility (see Clark & Flowerdew 1982 for a review). In the context of North America, Clark (1982a) and Huff (1982; 1986) have determined that housing searches are relatively short-term events.

Residential preferences

In addition to residential stress as a 'push' factor in household mobility, we can identify various 'pull' factors that draw households to certain submarkets and properties in a housing search. The hierarchy of preferences has already been discussed above. Much work has been done, researching household preferences for housing at both submarket and real estate scales.

In terms of price, households are understood to have a particular value-orientation (Golledge & Stimson 1997), and, of course, mobility is constrained to a very great degree by household income (O'Flaherty 1996).

Preferences for particular ethnic and social characteristics of residential neighborhoods have also been identified as being instrumental in explaining housing searches and mobility decisions. Models developed by Schelling (1971) and Sakoda (1971) explore these ideas in a general sense. Benenson and colleagues have also analyzed the nature and consequences of those forms of preference in empirical contexts (Benenson & Omer 2000b).

Tenure preferences are also important in residential mobility, and this may be linked to life cycle stages, as well as different cultural backgrounds (Benenson *et al.* 2002) and may be specific to particular locations.

Benenson and colleagues have conducted extensive analysis of household preferences for particular housing types and styles, and the role that those preferences play in household mobility and the evolution of residential neighborhoods (Benenson 1998, 1999; Benenson & Omer 2000a, b; Benenson *et al.* 2002). The conversion of preference to decision and choice heuristics is commonly specified using the utility models discussed in Chapter 2. Other techniques, such as "satisficing" (Simon 1956) and Markov chains (Huff 1982), have also been employed.

Specifying a GAS model of local residential mobility

The model described in this chapter was constructed with much of this theory in mind. Many components of the literature discussed above feature in the model design" hierarchy in housing search, stress and resistance, household life-cycles, and household preferences for submarkets and real estate. However, considerations above the level of the submarket are not treated; the model proceeds on the assumption that this information has come from elsewhere; in the context of this example, hypothetically, from the model discussed in the previous chapter.

The model is designed to represent households and properties in a very realistic fashion, with life-like characteristics, preferences, and mechanisms. Residential mobility is treated as a two-stage process. First, households decide whether to initiate a relocation or not, looking to both internal and external 'stressors' in formulating this decision. If they decide to engage in a housing search, they do so based on preferences that relate to their internal characteristics and those of the larger community/submarket in which they focus that search.

In the simulations described in this chapter, a single community/submarket is represented, comprised of many properties and households. The community/submarket is, itself, treated as an *entity* in the model, with the capacity to evolve dynamically as a function of the objects within it and to serve as an object that households actually interact within and with.

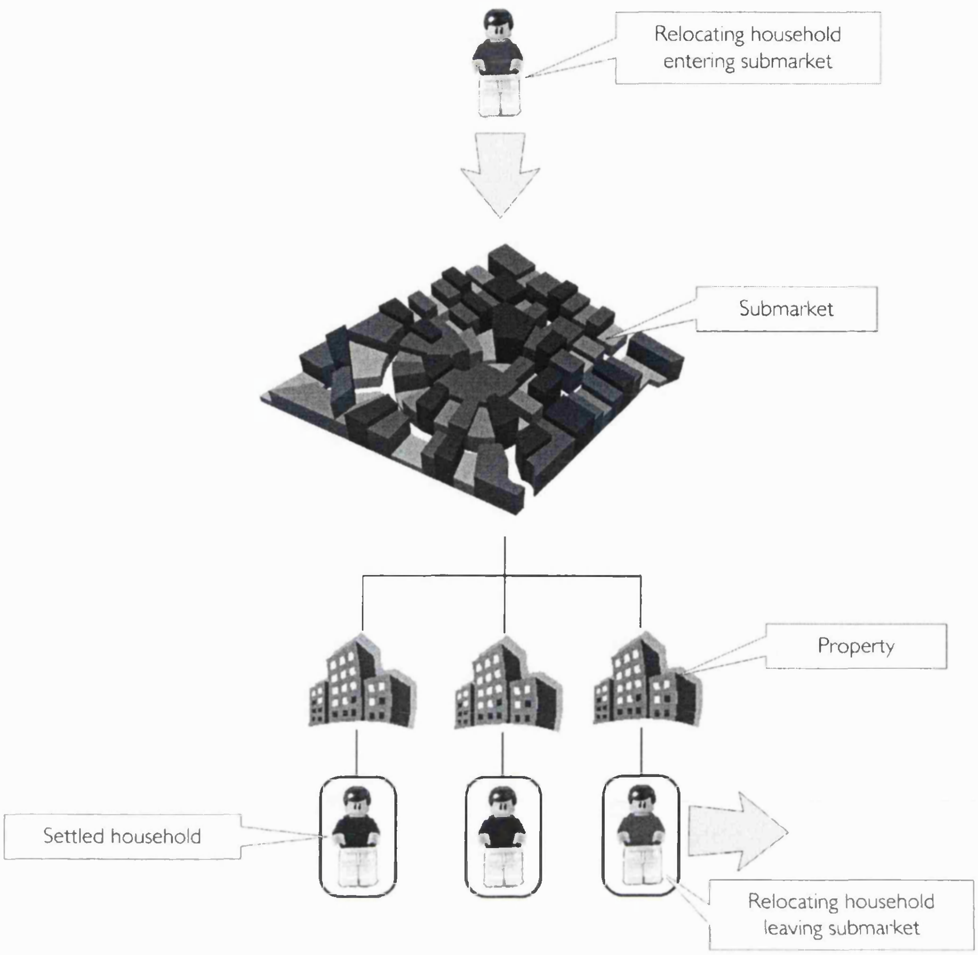


Figure 56. The hierarchy of entities in the residential location model

As before, the residential location model is specified as a GAS—a collection of GA designed with reference to typology, state variables, state transition rules, geo-referencing conventions, movement rules, neighborhoods, and neighborhood rules, and was programmed in StarLogoT 2001 (Center for Connected Learning and Computer-Based Modeling 2001). In addition, constraints are introduced to confine simulations within reasonable bounds.

Typologies

As was the case in the model described in the previous chapter, the residential location model is specified with two types of GA entities—fixed and non-fixed. In this example, the entities described are designed to represent different (nested) spatial scales in the model (representing Speare's (1975) taxonomy of factors governing residential mobility): sub-markets, the properties that comprise a sub-market (both of these are represented as fixed GA), and the households within properties (represented as non-fixed GA) (Figure 56). Sub-market and property GA do not move, although their internal conditions are usually quite dynamic in simulation. Household GA do move—into and out of submarkets, and within submarkets.

State variables

In the model described in the last chapter, GA were characterized as abstract units of the urban system. In this example, by contrast, GA are used to represent "atomic" (people-scale), spatially non-modifiable entities, and these entities are ascribed life-like attributes and behaviors that closely match the characteristics and functionality of their real-world counterparts.

In the model, submarket GA are simple aggregations of the property GA and household GA within them; because they are continually in a state of flux, the attributes of a submarket GA (Table 9) are dynamic in time.

Table 9. State variables for residential submarkets

State variable	**Symbol**
Total number of households	H
Total number of yellow households	E_y
Total number of blue households	E_b
Total number of red households	E_r
Median household income	I
Median household age	A
Median property value	v_{avg}
Maximum property value	v_{max}
Minimum property value	v_{min}

Table 10. State variables for fixed property automata

State variable	Symbol
Housing type (house, apartment)	h
Occupation status (occupied, not occupied)	o
Housing tenure (rent, own)	r
Sale or rental status (for sale, not for sale, for rent, not for rent, under offer)	s
Monthly mortgage value or rental value	v
Land-use (residential, non-residential)	u
Lot size	l
Density	d
Number of bedrooms	b
Time on the market	m

Table 11. State variables for mobile household automata

State variable	Symbol
Household type (settled, relocating)	y
Median monthly household income	i
Median household age	a
Number of children	c
Household size	z
Ethnicity (yellow, blue, red)	e
Lifecycle stage (young, middle, senior)	L
Period of residency, in time steps	p

Fixed GA are used to simulate the urban fabric; each GA corresponding to a piece of real estate. These property GA are designed with an array of authentic state variables, lending them property-like attributes (Table 10). The goal is to provide a 'smart', 'tactile' environment for household GA—a spatial landscape of socioeconomic attributes with which simulated households can engage and within which they can interact. In the current implementation, the model is specified with hypothetical real estate units. These are arranged in a 250-by-250-unit lattice of square-shaped GA, although only a handful of the units are activated at any given time. Residential real estate is represented, but other infrastructures are not.

Mobile automata are used to represent household units. Two classes of households feature: *settled* and *relocating* (the 'movers and stayers' that Brown and Moore (1970) write about). Settled automata correspond to households that occupy a property at any given instance in a model run. (Settled automata households are, in fact, static. However, they have the potential to move, so they are classed as mobile.) Relocating automata are synonymous with households that are actively searching for a new home.

Both classes of automata are equipped with life-like state variables that enable the specification of realistic household profiles (Table 11). A proxy variable is used to represent ethnicity (e), with households represented as either red, yellow, or blue automata. Following the discussion of life courses in the opening sections of this chapter, households are classified into particular lifecycle stages (L). This is a simplified interpretation of the lifecycle notion, but one that could be extended in principle. 'Young' households represent those that have recently left a family household unit and are striking out on their own for the first time. 'Middle' households are used to represent households that might be starting their own families or may have already started a family. 'Senior' households correspond to those households that are entering (or have already entered) retirement. Lifecycle stages are calculated based on the median age of a household (a):

$$IF\ 22 \le a \le 35,\ l = '\mathit{young}' \qquad \text{Eq xliii,}$$

$$IF\ 35 < a < 65,\ l = '\mathit{middle}' \qquad \text{Eq xliv,}$$

$IF\ a \geq 65,\ l = 'senior'$ Eq xlv

(Households with median ages below 22 are not considered in the model.)

State transition rules

Static property automata are roughly equivalent to CA, but differ in some important respects. As in the case of the urban growth model, static automata transition is mediated, for the most part, by mobile automata. In this example, property automata are equipped with a single transition rule of their own. If a property has spent too many days on the market (m) without selling or renting—above a certain threshold of days (θ)—then its mortgage or rental value (v) is discounted by a proportional value (λ) in a subsequent time-step, t+1. (In some senses, this could be interpreted as catering to supply-side factors, although they are assumed to have been generated exogenously from this model.)

$IFF\ m \geq \theta, v_{t+1} = \lambda v_t$, where $0 < \lambda < 1$, Eq xlvi,

otherwise $v_{t+1} = v_t$ Eq xlvii

In addition to possessing a set of life-like state descriptors, household GA are equipped with a set of preference functions that reflect those discussed in the residential mobility literature, and these serve as their transition rules in much the same way as in the last model. Preferences provide the connection between fixed property GA and non-fixed household GA and serve as the main rules driving mobile automata transition in the model. Mobile agents roam their simulated city environment, initiating state transitions in the fixed GA that they encounter—in this case, property GA. Because entities are nested in the model, transition changes in the property GA comprising a submarket GA register as submarket-level changes at that level, as conditions at a local-scale evolve. Household GA preferences are formulated as a set of 'likes' and 'dislikes', used to mimic the *behaviors* of households as they make residential location decisions in urban spaces.

Preference functions

Household GA exercise their preferences at two spatial scales: the residential submarket and the property scale (Figure 57). (This reflects Speare's (1975) ideas about search hierarchy as well as those of Clark (1993) and Huff (1986) in regard to area-based searches.) The implication is that households evaluate micro- (properties) and meso-conditions (submarkets) in their residential location behavior. Household GA evaluate an *inertia* preference at both scales, deciding whether they would like to relocate or not. This serves as their 'resistance' function. Resistance is specified in terms of environmental contexts—at the submarket scale, they act based on preferences for *segregation* and *wealth*. Within submarkets, household GA evaluate the suitability of real estate using *property* preferences.

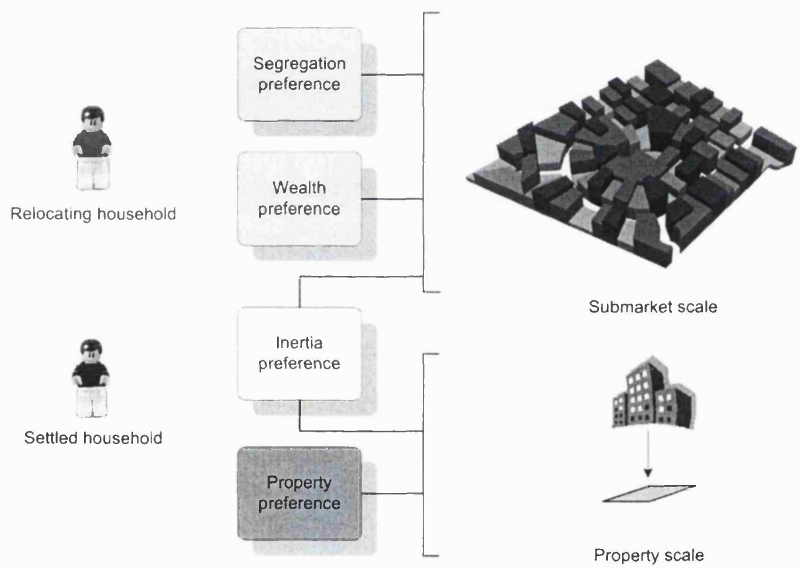


Figure 57. The organization of preferences across scales

A segregation preference function is used to represent households' likes and dislikes regarding the ethnic profile of residential submarkets. The function is specified in much the same way as Schelling's segregation models (Schelling 1978), using tipping points to specify households' comfort with certain conditions in the submarket, in this case ethnicity. In the simulation discussed in this chapter, yellow household GA have no preference regarding the ethnic composition of a submarket. Blue household GA exhibit a level of bias in their preferences; they have a preference for submarkets where red household GA form no more than one third of the total population. Red household GA prefer submarkets where they form a majority, above one half of the submarket's residential population.

Wealth preferences allow household GA to assess whether a neighborhood is too poor for them to enter or remain in, or whether it is too expensive to consider relocating to. This is equivalent to the value-orientation discussed by Golledge and Stimson (Golledge & Stimson 1997), among others. A submarket GA is regarded as too expensive for a particular household GA if the median monthly mortgage or rental value of a home (v_{avg}) is greater than the household GA's average income. If the maximum price in the submarket (v_{max}) is below a given household GA's monthly median income (i), then the submarket is regarded as being too cheap; if a household's income is lower than the lowest price (v_{min}), the submarket is too expensive.

In addition to preference functions describing household GAs likes and dislikes regarding submarkets of a given character, the model also provides functionality for detailing households' preferences for individual properties. Household GA have a preference for housing type (h) and housing tenure (r) (see Golledge and Stimson (1997) and Clark (1982b) for theoretical background). In addition, they have a preference for housing price (v). Type and tenure preferences are formulated, simply, based on the lifecycle state (L) of a household GA. 'Young' households have a preference to rent apartments, 'middle' households prefer to buy apartments, and 'senior' households prefer to buy houses. The expression of a housing value preference is also specified simply; households will not buy or rent a property that is more expensive than one third of their monthly income.

Geo-referencing conventions

A variety of geo-referencing conventions are used in the model. At the top level of hierarchy for entities in the model (Figure 56), submarkets are geo-referenced in terms of their position within a larger urban system. In the simulation described in this chapter, we consider a single submarket. But, multiple submarkets could be geo-referenced in a city-system by their absolute location within a lattice of submarket GA, each perhaps associated with some distance and/or accessibility to a city center or centers.

Property GA are geo-referenced directly, using the X, Y coordinate location of their centroid within the property GA lattice. Household GA are then associated with this point, thereby avoiding conflicts in terms of 'agents' occupying single 'cells'.

Household GA are geo-referenced both directly and indirectly. Settled household GA are geo-referenced to their property using X, Y coordinates. Relocating household GA are geo-referenced directly—their absolute position within the lattice is registered using X, Y coordinates. As they search for potential homes, they are geo-referenced indirectly with respect to the last property they viewed and the next property that they plan to view.

Movement rules

Movement is specified in the model in terms of the entry and exit of households to and from submarkets, and the occupation and displacement of households in and from properties. This is migratory movement, rather than locomotion; non-fixed household GA iterate between various property options, evaluating each as they encounter them.

Neighborhoods

Neighborhood definitions also differ from those specified in the last model. Neighborhoods for submarket GA consist of all household and property GA active within a submarket. A variety of statistics are generated from these neighborhoods, dynamically, when a simulation is run. The values register as state variables in submarket GA (Table 9).

For relocating household GA, the neighborhood initially consists of a submarket when they first enter the simulation. Once they begin to search a submarket, their neighborhood consists of the property GA that they chose to evaluate. At a given transition point after that they register their current location (which may be a property GA), previously-viewed property, and next-view property as their neighborhood (Figure 56).

Constraints

Simulation runs are constrained quite simply—a volume of relocating households is delivered to the model at the start of a simulation run. Similarly, a number of fixed property GA are active on the housing market at the start of a particular simulation run.

Dependencies between state variables and preference functions also create a set of constraints. Relocating household GA may only purchase properties they can afford. Also, relocating households will only consider submarkets that are either affordable, or above a certain price threshold. The latter ensures that relatively more affluent households do not relocate to relatively poorer-profile neighborhoods. Furthermore, households are constrained to searches for properties that have enough bedrooms to accommodate their household size.

A typical simulation run

Model runs are organized as a series of events, roughly designed to coincide with the various stages that constitute residential mobility decisions (Clark 1982b). Several sub-events take place in parallel within the model, such as calculations and the derivation of preference functions, but the main events that drive a model run occur iteratively (Figure 58).

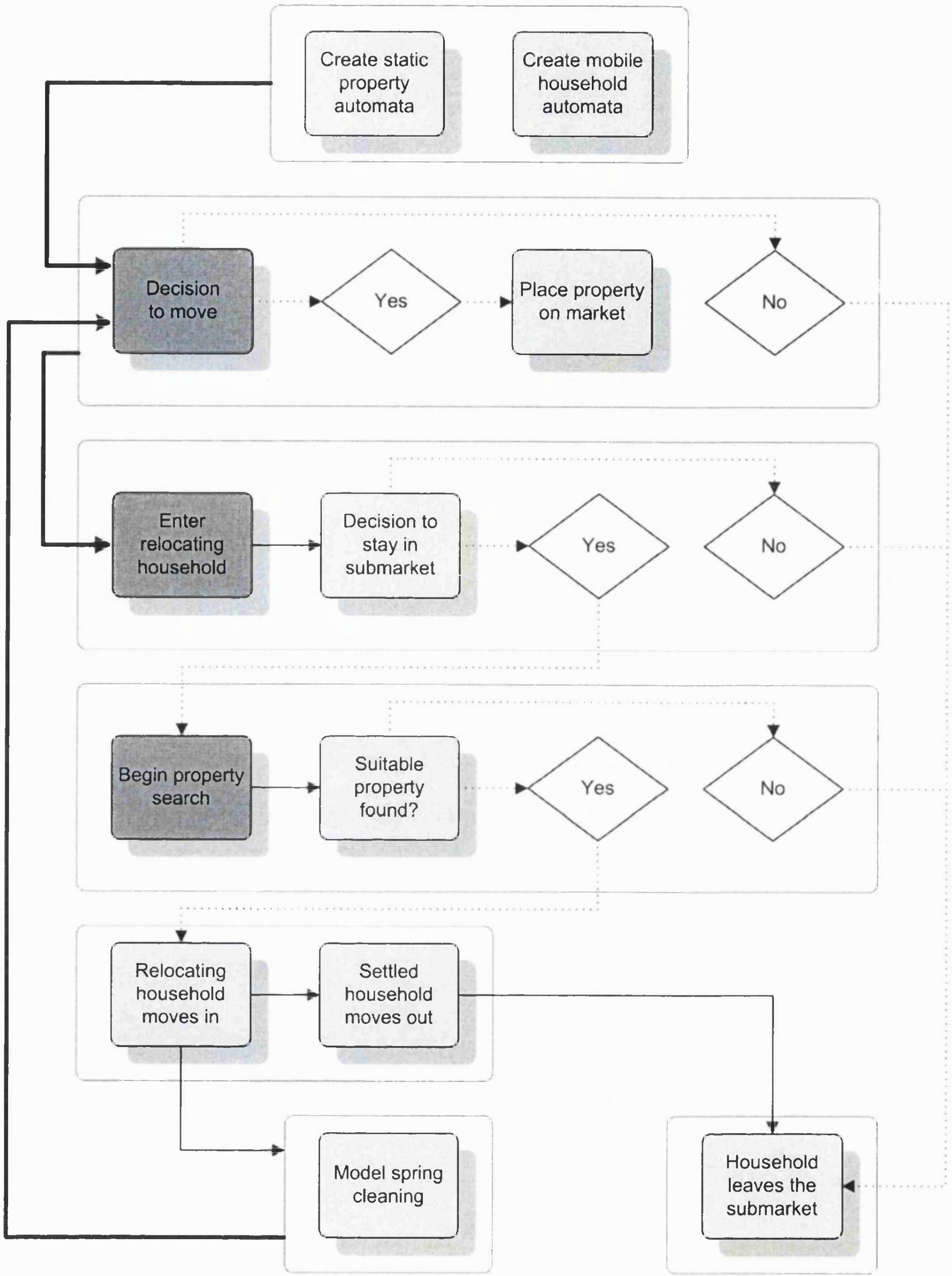


Figure 58. The sequence of events in a residential location model run

The first stage in a model run involves the creation of a virtual residential environment. 'Robot' GA are called upon to seed the model with static automata properties. These robot GA enter the model and code attributes into a blank lattice-space, establishing a simulated residential property market. Specific units are rendered active in the model and are assigned property attributes (Table 10). Following this, activated property GA are then populated with settled household GA, characterized with life-like attributes (Table 11).

Once individual property sites have been established and populated, the settled household GA that occupy them make a decision whether to move home or not. This is a multi-scale evaluation: household GA look both to their internal household characteristics and also to the attributes of the submarket in which they reside before making a decision. Changes in households' internal states (internal stress) may result in a decision to move, e.g., transition to a new lifecycle stage, growth or reduction in household size, the arrival of children, etc. At the same time, there may be forces at work at the scale of the submarket (environmental stress) that influence their decisions to relocate or not. These can initiate endogenously within the system, e.g., changes in the socioeconomic profile of a neighborhood through gentrification or neighborhood decline, or could be specified as exogenous shocks such as the construction of new highway. If a settled household GA decides to move, its property (the individual fixed property GA that the household is associated with) is flagged as being for sale or for rent, as the case may be. This registers in the appropriate state variable for the GA.

At this stage, a relocating household GA enters the model, in search of a new home. Currently, incoming GA are generated synthetically, from the top-down, and enter the simulation as a simple feed. Relocating GA hold a set of preferences for their ideal location and home, and must balance these desires within the bounds of what they can afford (resistance). Currently, only one relocating household GA populates a given model run at any instance, although many settled automata are available for it to interact with. First, a relocating household GA looks to the submarket to determine if it is suitable for its needs—whether it has the right 'ethnic' profile and whether it is too expensive or too cheap. If the submarket is suitable, the relocating household will begin to focus its housing search on individual properties in the submarket. Potential homes are assessed for their suitability to individual households' needs.

The relocating household GA visits active properties randomly, evaluating their characteristics against its preferences. In the simulation described in this section, the evaluation is formulated in a hierarchical fashion. A relocating GA begins the evaluation by checking that the housing type (h) matches that in its preference set. Next, the relocating household GA evaluates tenure (r), followed by price (v). This is quite a simplistic method for matching preferences with attributes, although it does provide for the weighting of choices in the tradition of the discrete choice models mentioned in Chapter 2.

If a relocating household GA finds a suitable property that matches its preferences and its budget, it trades places with the settled household that occupies the property. The sale (or rental) status of the property is updated to reflect the fact that it is no longer on the market. The new household is moved in, changing status from relocating to settled, and the original household is moved out, again changing status. These adjustments are also registered in the global characteristics of the submarket, thereby facilitating neighborhood-wide changes. For example, if a red household replaces a blue household, the balance of colors in the neighborhood will be altered to reflect that; if a household with a very high median income decides to move into a neighborhood, the maximum value for median income in the neighborhood will be revised to indicate the change.

The model is then put through a series of "spring-cleaning" exercises before beginning a new iteration. Any relocating GA that have not satisfied their searches after visiting all available properties leave the submarket (that is, their 'stopping rule' is triggered by the failure of the search iteration in a submarket). If settled households have not sold or rented their property, they may decide to discount the price they are offering the property for in the next iteration of the model. By iterating through these events in a sequential fashion, the model allows for the evolution of the submarket, its population, and the individual properties contained within it. Submarkets could, potentially, go through cycles of decline and gentrification, for example. The socioeconomic composition of the population could also be allowed to cycle through various stages, e.g., from a youthful profile to one more characteristic of empty-nesters. The introduction of shocks of various descriptions to the submarket could also allow for the exploration of submarket responses to things such as an influx of wealthy households, from the bottom-up.

Experimenting with the simulation

A simulation was run with the intention of evaluating the impact of introducing households of various descriptions to the model, and the effect it would have on the dynamics of the submarket. Essentially, the goal of the simulations is to test the resilience of the systems to change from the bottom-up. Evaluating the impact of incoming households on the price profile (value platform) and ethnic make-up of the submarket was a particular focus. This is quite relevant to submarkets in sprawling areas of American cities, where exclusionary practices can take place, either deliberately or through the collective but independent preferences of resident households.

An artificial submarket was created, consisting of a 250-by-250-unit lattice, in which 29 properties were active on the market (29 settled household GA were seeking to sell or rent their homes). The settled households were specified randomly in terms of their ethnicity and median income. This generated a submarket with ten red households, six yellow households, and nine blue households (Figure 59).

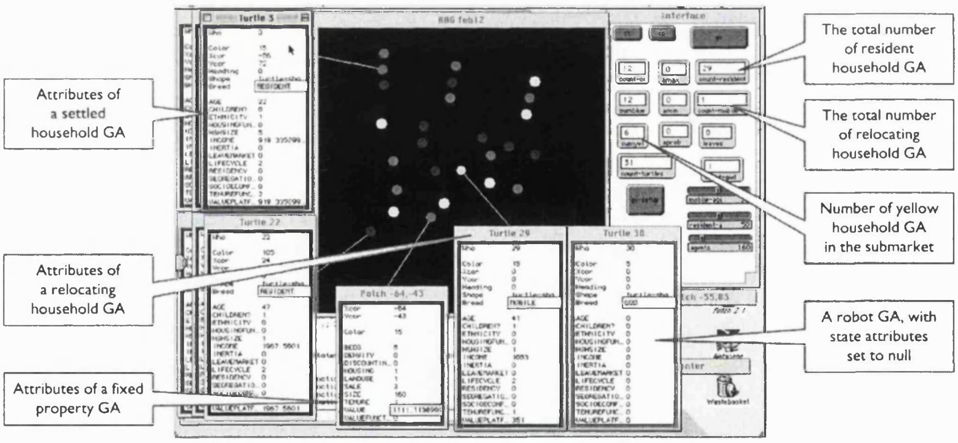


Figure 59. The residential model GUI

Specifically, mobile household automata were assigned attributes on a random basis; fixed property automata were specified as follows:

Site location	Housing type	Occupation status	Housing tenure	Sale/rental status	Monthly rent/mortgage	Land-use	Lot size (m^2)	Density	Bedrooms	Days on market
(-55, 83)	Apt.	Occ.	Rent	Occ.	3000	Res.	100	0.01	2	0
(-5, 74)	Apt.	Occ.	Rent	Occ.	3300	Res.	120	0.0083	2	0
(52, 63)	Apt.	Occ.	Rent	Occ.	3500	Res.	140	0.0071	1	0
(-56, 72)	Apt.	Occ.	Rent	Occ.	3000	Res.	90	0.0111	2	0
(-7, 62)	Apt.	Occ.	Rent	Occ.	3000	Res.	100	0.01	2	0
(48, 51)	Apt.	Occ.	Rent	Occ.	3400	Res.	110	0.0091	2	0
(-55, 60)	Apt.	Occ.	Rent	Occ.	5000	Res.	130	0.0077	2	0
(-9, 51)	Apt.	Occ.	Rent	Occ.	5200	Res.	140	0.0071	2	0
(46, 40)	Apt.	Occ.	Rent	Occ.	6200	Res.	160	0.0063	3	0
(-57, 33)	Hse.	Occ.	Own	Occ.	3000	Res.	100	0.01	4	0
(-27, 31)	Hse.	Occ.	Own	Occ.	4300	Res.	110	0.0091	3	0
(15, 23)	Hse.	Occ.	Own	Occ.	3000	Res.	85	0.0118	4	0
(34, 18)	Hse.	Occ.	Own	Occ.	3000	Res.	80	0.0125	2	0
(52, 11)	Hse.	Occ.	Own	Occ.	5000	Res.	100	0.01	4	0
(-27, 19)	Hse.	Occ.	Own	Occ.	3000	Res.	90	0.0111	3	0
(21, 10)	Hse.	Occ.	Own	Occ.	2500	Res.	90	0.0111	4	0
(-48, 6)	Hse.	Occ.	Own	Occ.	3000	Res.	200	0.005	4	0
(-18, 1)	Hse.	Occ.	Own	Occ.	5500	Res.	210	0.0048	3	0
(20, 19)	Hse.	Occ.	Own	Occ.	5000	Res.	175	0.0057	3	0
(40, 14)	Hse.	Occ.	Own	Occ.	5000	Res.	180	0.0056	4	0

Site location	Housing type	Occupation status	Housing tenure	Sale/rental status	Monthly rent/mortgage	Land-use	Lot size (m^2)	Density	Bedrooms	Days on market
(46, -14)	Hse.	Occ.	Rent	Occ.	2000	Res.	50	0.02	5	0
(-12, 14)	Hse.	Occ.	Rent	Occ.	5500	Res.	100	0.01	2	0
(24, -44)	Hse.	Occ.	Own	Occ.	6500	Res.	400	0.0025	3	0
(73, -65)	Hse.	Occ.	Own	Occ.	7500	Res.	400	0.0025	6	0
(-21, -33)	Hse.	Occ.	Own	Occ.	6500	Res.	110	0.0091	4	0
(-48, -30)	Hse.	Occ.	Own	Occ.	10000	Res.	350	0.0029	3	0
(-64, -43)	Hse,	Occ.	Rent	Occ.	4000	Res.	160	0.0063	5	0
(13, -72)	Hse.	Occ.	Own	Occ.	5600	Res.	100	0.01	3	0
(-55, -76)	Hse.	Occ.	Own	Occ.	6000	Res.	550	0.0018	6	0

Once the artificial submarket was established, relocating households were introduced to the submarket, and set to search for potential homes. To test the reliability of the model, households with very low median incomes were introduced, as were households with very high median incomes. Both evaluated the income profile of the submarket and promptly terminated their search, as their preference rules for value platform determined. Next, a household with a price preference that was within an acceptable range was introduced. The household elected to begin a search within the submarket. However, its preference for housing type was not satisfied within that price-range, and it left the submarket after evaluating all available properties.

A series of targeted experiments were then performed with the model. A stream of rich households of varying colors was introduced to the simulation, with incomes within the acceptable bounds for property prices in the submarket. As relocating households found suitable properties to buy or rent, they moved into the submarket, displacing the previous residents. If the incoming household displaced a household of a different color, the ethnic profile of the submarket altered to reflect that change. Slowly, the balance of ethnicity in the submarket began to change. This prompted two effects. The incoming households began to react differently, rejecting the submarket depending on their particular preference for ethnic profile. Also, the internal dynamics of the submarket began to change. As the ethnic balance of the submarket altered, the 'stressors' of previously-settled household GA were *triggered* by the shift in balance, and they began to place homes on the market. By varying the ethnicity of incoming relocating household GA, it was possible to shift the ethnic balance of the submarket from mixed conditions to dominance in each of the colors, and from segregated to mixed conditions.

Another experiment was run to investigate lifecycle dynamics in the model. Relocating households with lower-than average and higher-than-average median ages were introduced to the simulation. As these households began to occupy properties in the simulation, so the demographic profile of the submarket began to change accordingly. Similarly, experiments were performed whereby household size variables were changed for particular settled households. If the increase in household size surpassed the number of bedrooms in their home, they placed their property on the market for sale or rent.

A similar experiment with the socioeconomic profile of the submarket produced some interesting results. The discounting function, designed to lower the price of a property in the simulation—if there was insufficient demand for it—provided only minor changes in the economic dynamics of submarkets. This is because budgets for relocating households were based on their income in the model; this ensured that very affluent households did not buy or rent very cheap houses, and vice-versa. So, small reductions in property values did very little to permit access to submarkets for lower income groups, particularly because incoming relocating households looked to the *average* value of the submarket before deciding whether to continue a search. This is what happens in the real world; gentrification and decline are slow processes, and are likely to be more dependent on factors such as crime, social problems, and the quality of recreation and retail opportunities in the submarket—factors not represented in this model. However, the price profile of the submarket was also quite resilient to changes in the *internal* dynamics within the households that form the submarket. Adjusting the median income of individual settled households did not affect the price of property in the region, because the price at which a household sold or rented a property was not linked to its income in the model; at this scale, price was purely demand-driven.

The resilience of the model to these changes is consistent with situations in residential submarkets in many sprawling areas in the United States. The sorts of phenomena observed in these experiments are indicative of lock-out situations in wealthy suburbs, whereby lower income groups are barred entry to suburban areas because properties there are not affordable (see Cronin 1982; Farley & Frey 1994; Farley *et al.* 1978; Galster 1991; Knox 1989 for discussion in American contexts). Even if affordable housing is constructed in these areas, lower income groups may still be excluded by minimum lot requirements—zoning codes that place a minimum size on residential lots. For example, a minimum lot ordinance calling for half acre lots raises the overall price of real estate, even if the structure price of a property is relatively low, simply because the land component of the price remains high. The results reported for these experiments suggest that suburban submarkets, where property prices may well be uniformly high, are very resilient to change in that price platform.

Conclusions

In Chapter 8, a distribution-based model for exploring geographic determinants of sprawl was described. That model was specified around a GAS core, and designed to operate at a synoptic scale—at the level of a region-wide city system. The smallest units in the growth-based simulation were landscape GA, each representative of a large area of landscape or urbanized land, roughly equivalent to the size of a residential property submarket.

The model introduced in this chapter extends that idea, conceptually, focusing on very local-scale dynamics, as they might take place *within* a single landscape GA unit from the last model. The model described in this chapter focuses on residential location dynamics, specified at an "atomic" scale—describing individual households and properties, dynamically active within a larger residential submarket.

The switch to a local scale continues within the model parameters, as well as manifesting in the fidelity of entities within the model. The residential location model is built around a GAS framework. Individual-scale entities are described with life-like state variables. However, their independent decision-making behaviors are also specified, with transition rules designed to represent preferences for housing (type, price, tenure), and the evaluation of those preferences with respect to individual properties and submarkets in a simulation. In particular, preferences for submarkets of certain ethnic, demographic, and socio-economic profiles are supported.

A series of experiments run with the model were described in this chapter, focused on evaluating the sensitivity of the simulated community/submarket to changes in its economic, demographic, and ethnic profile, as evolved from the bottom-up. The experiments demonstrated that the model was quite resilient to changes, and this is indicative of the situation in many submarkets in sprawling areas of the United States. Several conclusions can be drawn, first in relation to the suitability of the GAS framework for analyzing urban dynamics, and second, with respect to suburban sprawl.

The GAS framework described in Chapter 5 worked well in the context of the experiments documented in this chapter. It thus appears to be quite applicable in multiple scenarios, and at multiple spatial scales. In particular, the spatial functionality of the framework facilitated the design of simulations that closely represent residential property markets and their dynamics. Entities of varying sizes could be designed, and nested neatly with respect to their position in a hierarchy. Varying neighborhood functions were also facilitated, as were direct and indirect geo-referencing conventions. Also, the ability to accommodate migration rules was also quite significant.

The experiments described in this chapter offer some insight into sprawl issues, as they might evolve very locally within a city system. It is clear that local-scale dynamics "within the cell" are very important, particularly in explaining the price dynamics of housing, as well as some of the socioeconomic and demographic determinants of sprawl at a local scale. The GAS framework, with its ability to describe independent units, is quite useful in evaluating these phenomena. However, higher-scale dynamics, as explored in the model in Chapter 8, are as important in the context of sprawl. The conceptual linkage of the two models in this thesis points to potential benefits for adopting a multi-scale approach to looking at the dynamics of sprawl.

To a certain degree the approach demonstrated in this chapter also breathes new life into the residential mobility literature described at the start of this chapter. The infusion of spatially-explicit functionality and mechanisms, in particular is significant.

Chapter 10. Conclusions

"I pointed out that his copy of *Binary File Transfer Monthly* was possibly the most boring document I had ever seen in my life." (Coupland 1995, p. 167)

Introduction

This chapter draws the thesis to a close, synthesizing the discussion in the other nine chapters and drawing conclusions from the research exercise. This section of the chapter discusses the objectives of the thesis research and reviews the approaches taken to realize those goals. The following section examines the results of the research in terms of simulation as a methodology and technology for (urban) geographic analysis, and sprawl as a spatial phenomenon. Finally, the chapter concludes with discussion of future avenues of exploration for this work.

In the opening lines of Chapter 1, this thesis started by asserting its focus in two areas—models and sprawl. *Space* has been the dominant theme through which this focus has been investigated. In particular, the discussion in the text, and the research behind it, focuses on embedding space into simulation technology, applying that technology to the study of geographical systems (suburban sprawl in particular), and using the technology to explore the geography of change in simulated cities.

In Chapter 1, the goals of the research were identified with two main aims. First, to develop explicitly *spatial* simulation technology to support research into urban dynamics. Second, to use that technology to explore the geography of suburban sprawl in a North American context. These objectives were approached in a number of ways.

A new framework for *geographic* simulation at fine resolutions and in complex, dynamic contexts was developed: Geographic Automata Systems (GAS). This framework extends already popular simulation methodology from the computing sciences—basic automata, cellular automata, and multi-agent systems. Additionally, the GAS framework introduces some core, geographic, functionality that offers unique advantages for spatial simulation.

The framework is demonstrated, operationally, through application to the modeling of suburban sprawl in a North American context. Sprawl has some particularly spatial characteristics and mechanisms and provides an excellent domain for simulation in a GAS framework. Two classes of model were built, each with a different perspective on the sprawl phenomenon.

The first model was developed to simulate sprawl from a spatial distribution perspective. Simulations with the distribution-based model start with initial conditions of population growth in a city-system, and focus on the spatial distribution of that growth to local-scale geographies, but with an emphasis on the spatial *mechanisms* driving growth. Three simulations were developed. Two abstract simulations were run; the focus was on the impact of variable rates of change in the system and on the relative significance of the spatial mechanisms of distribution in the model. A third simulation was developed to simulate the dynamics of sprawl in the Midwestern Megalopolis city-system. The emphasis, in that simulation, is on the impact of growth rates and the comparative impact of the spatial factors of distribution in the simulation, as before. However, the Midwestern simulation was constrained with known conditions about the system of interest, thereby calibrating it to real world conditions over a historical period. In addition, techniques from spatial analysis were used to assess model output, empirically.

The second class of GAS model approached the sprawl phenomenon from an altogether different perspective. The model described in Chapter 9 was developed with very local-scale attributes of sprawl in mind, focusing on urban dynamics within a residential submarket, and among the community of households—and their properties—that comprise it. In terms of spatial mechanisms, the focus was on preferences for real estate and urban living, the decisions that households make when searching for a home, and the expression of those decisions and preferences in a community context. The model was designed to explore individual- and community-level dynamics that might have implications for sprawl formation. A series of experiments were performed with the model, testing community-wide response to economic, social, and demographic change, introduced to the system from the bottom-up.

The thesis innovates in a few ways. In terms of simulation *technology*, the GAS framework appears to be very useful in supporting geographic simulation. It facilitates the union of cellular automata and multi-agent systems concepts, with *geography* as the binding force. The framework is very resilient to implementation across a range of different scenarios, as demonstrated in this thesis. It also allows for the expression (and exploration) of uniquely geographical concepts in a simulation; the implications for examining sprawl formation will be discussed shortly.

In terms of understanding *sprawl*, the research documented in this thesis makes several contributions. The magnitude of the sprawl phenomenon, and seriousness of its apparent consequences, are such that ideas and hypotheses are required to inform the debate, in academic circles, but also in planning and policy contexts. However, phenomena like sprawl do not lend themselves to experimentation in the real world. There is relatively little utility in throwing proposed solutions at the phenomenon and hoping that they will work. Simulation has a role to play in informing the debate, to enable the creation of virtual laboratories to support what-if scenarios to evaluate the likely or potential impacts of action on the phenomenon, but virtually, before acting in the real world. However, sprawl is illustrative of new kinds of phenomena—or even new understanding of old phenomena—in urban systems. These are complex systems, derived from the interactions of many diverse agents, mechanisms, and sub-systems, manifest at different temporal and spatial scales. Ironically, to approach a simulation of sprawl, one requires a model that is as decentralized in its structure as the phenomenon it is emulating. The experiments described in this thesis have been designed to do that, and to approach sprawl from that perspective, focusing on the role of geography in sprawl formation.

Results

The research described in this thesis has generated results in two areas: simulation and sprawl. The remainder of this section focuses on results under these classifications.

Conclusions about spatial simulation

One of the results of this work has been verification of the significance of the automata concept in simulation. All of the advantages outlined in Chapter 4 were supported in the experiments performed in this research.

The level of spatial detail afforded by the automata approach was very significant. It allowed for the specification of *entity*-level components and mechanisms responsible for sprawl. The decentralized nature of automata was also useful. All of the geographic mechanisms in the models supported decentralization in unique ways. The models facilitated decentralization through space and time in a theoretically-intuitive fashion, and allowed for the simulation of sprawl as a decentralized and decentralizing phenomenon. The dynamic attributes of the automata approach were essential to the simulation exercises. The models described in this thesis are all highly dynamic, and dynamics was one of their core components. This allows for an emphasis on space-time dynamics in the simulation of sprawl formation, the importance of seed conditions in inertia, and path-dependence in the evolution of the synthetic sprawling systems. The ability of automata to capture properties of function and form is also important. The validation exercises in Chapter 8 demonstrate the close symbiosis between pattern and process in the sprawl simulations. The fact that the two could be linked, and tied to ideas about similar relationships in the real world, is significant in itself. Cellular automata and multi-agent systems support a great deal of geographical simulation functionality; GAS offer even more. The affinity between automata and geographic data structures also proved to be beneficial in the simulations. The raster base of Geographic Automata units in the simulations facilitated capture of the state of the simulated system at any point in a run, and dynamically across a run. As was demonstrated in Chapter 8, in particular, this facilitated interpretation of results directly. This point is closely related to the advantages that automata tools offer for visualization. The discussion in Chapter 8, in particular, relies heavily on visualization, with the model developer (me!) interpreting results through visual inspection and by running through simulation runs dynamically on the screen. Visualization also supported empirical validation; the actual visual output—maps—generated by the simulations were used as the basis for validation exercises. Finally, the benefits of analogies between automata and complexity studies also proved useful. The emphasis on interaction and bottom-up dynamics was especially significant for capturing a diverse range of attributes of sprawl. In the models described in Chapter 9, particularly, it offered a unique perspective on the system, allowing community-level attributes to be inferred from individual action, but without ecological fallacy.

The GAS framework was very significant in a simulation context. The ability to divide modeled entities into typologies provided a natural system for classifying modeled objects. It allowed hierarchies of entities to be established, distinguished by fixed or non-fixed status. This was fundamental to both models. The location conventions are extremely useful, allowing for the infusion of geography into a model and the simulation of behavior related to the location of entities in space. Modeled objects behave differently when moving or still, when in fixed space or non-fixed space. Essentially, the location conventions provide a framework for linking typology, movement rules, state transition rules, and neighborhood rules, allowing GA to shift between cellular automata conditions and multi-agent systems almost interchangeably. The movement rules from the GAS framework are absolutely essential to both models described in this thesis, and to the geographic characterization of sprawl as a spatial phenomenon. In the regional model, movement rules were designed to mimic the actual motion of development in the simulated system. The formulation of movement rules as migration in the residential mobility model demonstrated that non-movement—inertia—could be as significant as motion. It is notable that the movement rules actually explain almost all of the geographic functionality in both models, regulating fixed Geographic Automata to a container and diffusion mechanism. This is, perhaps, getting at the heart of human-environment interaction in systems like sprawl—the important thing in *human* systems is the actual humans. A lot of automata models abstract from this notion (Torrens & O'Sullivan 2001).The neighborhood rules from GAS were also significant. They serve as a mechanism for relating action and activity across entities at different scales in a simulation—developers and settlers and their gateways; households and properties and their communities and submarkets. Neighborhood rules are also a natural mechanism for relating (mobile and dynamic) objects in space and time, both directly and indirectly, and connecting them with the fixed infrastructure in which they may be animated. For example, households were related to neighbors, properties, submarkets, and communities, all at once, with all domains of interaction factoring into their behavior.

There are other implications of the research for simulation. *Geography matters!* Many of the tools described and utilized in this work have origins in computer science, but can be adapted, uniquely, to spatial purposes through the infusion of concepts from the geographical sciences. This includes mechanisms such as spatial analysis for analyzing simulation output. More theoretical concepts can also factor into model development, however, such as theories about humans and their environments, movement and relationships in time and space, neighborhoods of influence and the friction of their impact, etc.

Dynamics, *within the cell*, also matter. Urban systems are simultaneously more than abstract objects and more than mentalistic entities; they are closely-bound and symbiotic mixtures of both. The simulations described in this thesis demonstrate opportunities for exploring spatial dynamics within the cell and relating them to conditions outside of the cell. This sort of cross-scale analysis is very useful and suggests opportunities for re-visiting geography at micro- or "atomic-scale" levels of observation.

Conclusions about suburban sprawl

One of the main conclusions stemming from the work documented in this thesis is that sprawl is inevitable. It is a by-product of growth, and also of decentralization. These are conditions that are almost universal in North American cities. Of course, those conditions exist in other countries, particularly in developing nations, but they do not form sprawl (in some countries they actually lead to dense informal settlements). Other factors are thus important in explaining the phenomenon.

Geography is extremely important. The fact that the simulations in this thesis were able to generate visually and configurationally realistic synthetic sprawl with geographic mechanisms alone is itself indicative of the significance of spatial factors. Growth is the fundamental driver of sprawl, but the mechanisms that distribute that growth in space are crucial to understanding the phenomenon. The simulations described in Chapter 8 were used to test various alternatives and to determine the spatial mechanisms that might be important in governing sprawl.

This suggests a role for geographical concepts in managing sprawl. This is essentially what we examined in the simulations—we re-wrote the rules of how cities work and examined what would happen, city-wide. Encouraging polycentric development seems to be the key. A combination of polycentricity and *local-level* sprawl seems to be most useful in stemming an inevitable tide of sprawl. These results match the observations discussed—controversially—by authors at the University of Southern California (Gordon & Richardson 1997a, b; Peiser 1989), but also suggests that local-scale initiatives such as transit villages (Cervero 1998) and edge cities (Garreau 1992) may actually help to cement otherwise diffusive urbanization trends. The results of the simulation exercises also point to the pitfalls of road-building without a strategy for encouraging sustainable road-induced development.

Another finding from the research is that the local scale matters. Sprawl should, perhaps, be tackled at both macro- and micro-levels. The modeling exercise described in Chapter 9 suggests that, at a community scale, local areas are very resilient to change under certain conditions, particularly conditions generally observed on the sprawling periphery, even without a collective sense of resistance. This suggests that either change in behavior is necessary, or more likely, alternative residential communities should be encouraged. Of course, this is what the New Urbanists (Calthorpe *et al.* 2001; Duany *et al.* 2000; Duany *et al.* 2001) are trying to do, although arguably with limited success thus far.

Future directions

The research described in this thesis has fostered a number of ideas and promising threads for future work in this area. The continued development and application of the GAS framework is one avenue that offers promise for future inquiry. The author has been working closely with Itzhak Benenson at Tel Aviv University to continue advancing the GAS framework, as an idea and an operational simulation technology for application in studying urban dynamics (Benenson & Torrens 2003; Torrens & Benenson 2003). The role that geography plays in 'new wave' simulation is another keen interest (Benenson & Torrens 2004a, b; Torrens 2003a; Torrens & O'Sullivan 2000b, 2001), and a book with Itzhak Benenson is in preparation on the topic (Torrens & Benenson 2004).

Validating these models is another avenue for further study, and that goes hand-in-hand with work the author has engaged in, looking at ways to identify and measure sprawl, empirically, using spatial analysis (Torrens 2004b; Torrens & Alberti 2000) (Figure 60, Figure 61). Preliminary work has been carried out, investigating potential explanations of that structure, examining Austin in Texas very comprehensively over a fifteen-year period, and also looking at digital maps of virtual worlds (the output is very similar to that generated by GAS simulations in this thesis) (Shiode & Torrens 2003a, b)[9] (Figure 62).

Figure 60. Population density in America's Northeastern Megalopolis, 1990.

Exploration into more of the mechanisms of sprawl discussed in Chapter 6 is another desired future research direction. The work described in this thesis represents the supply-side factors of sprawl in an abstract—although geographical—way in Chapter 8, and linked them conceptually to decision-based demand-side factors in Chapter 9. Future work is planned, to construct a development model, again based on spatial decision-making, to compliment the residential location model. The idea is that multiple submarkets could be designed, evolving interactively at varying scales of spatial resolution. This is also a large-scale test for the GAS framework, but would be

[9] Winner of the Best International Planning paper award at the Sixth Sharjah Urban Planning Symposium, June, 2003.

potentially very useful in testing policies such as growth management, and their geographic dimensions from structural and behavioral perspectives.

A further interest in dataware for simulation has been fostered by this work, and reinforced through involvement with my co-editors on the Digital Earth project (http://www.digitalearth.org). This has seen the author publishing in topics outside of the particular field of inquiry discussed in this thesis (Torrens 2004b). I am taking this as a good sign!

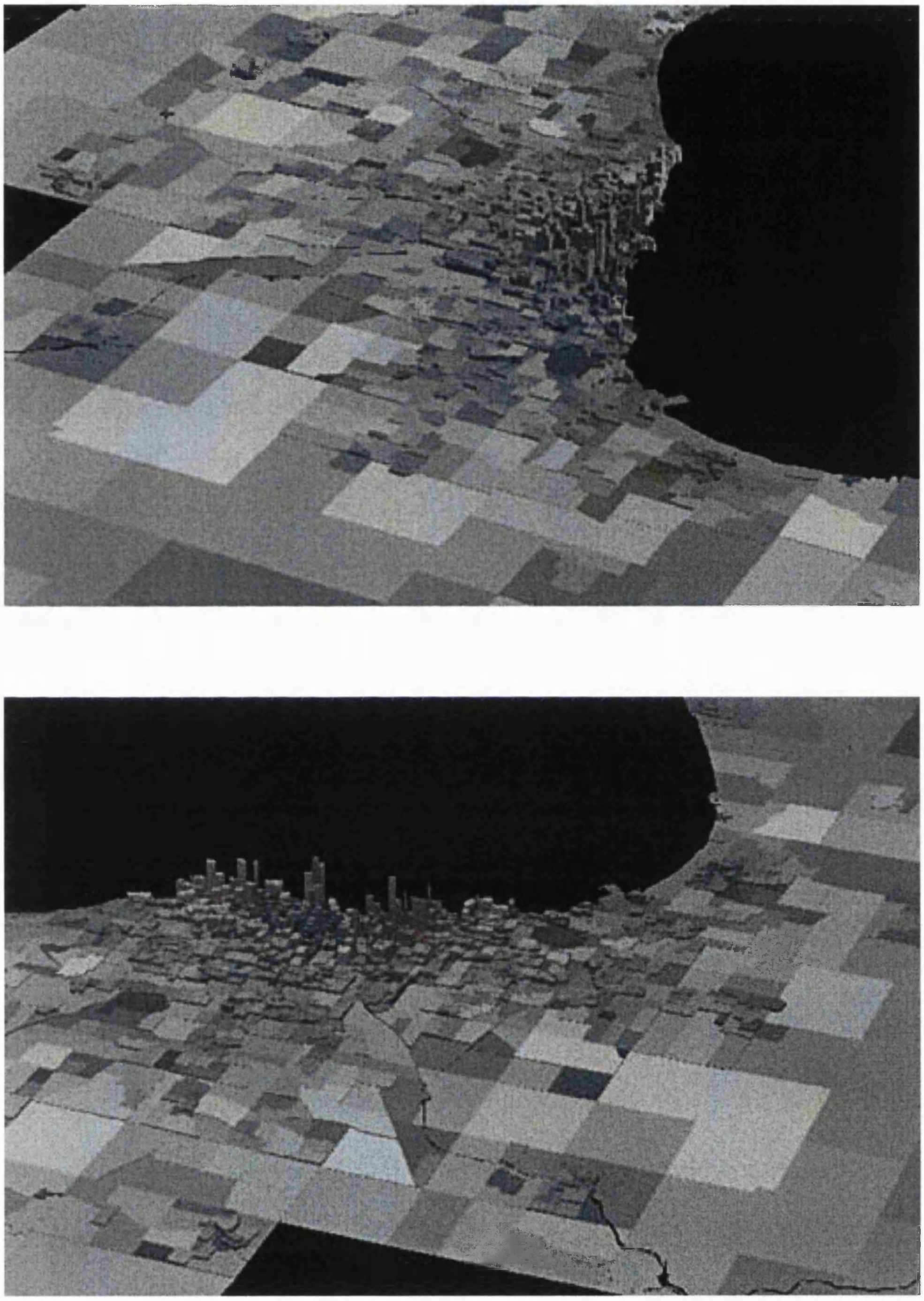
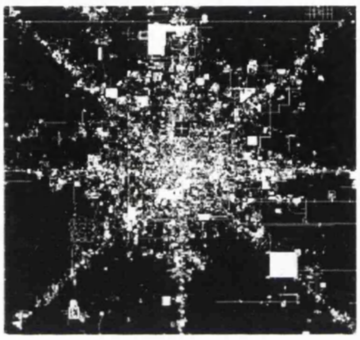


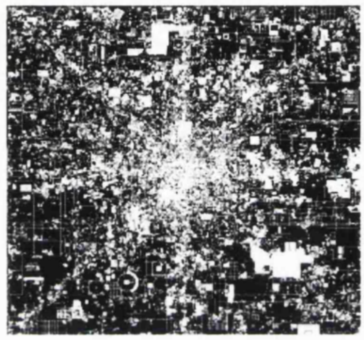
Figure 61. Three-dimensional visualization of population density in Chicago's Loop, 1990.

Urban extent (black), Austin, TX, 1990

Urban extent (black), Austin, TX, 1995 (growth shown in grey)



Urban extent (white), Alphaworld, 1996. (Source: Activeworlds.)



Urban extent (white), Alphaworld, 2001. (Source: Activeworlds.)

Figure 62. Visualization of urban growth in a real and virtual city.

References

Adams, D. 1994. *Urban Planning and the Development Process*. London: UCL Press.

Alberti, M. 2001. Quantifying the Urban Gradient: Linking Urban Planning and Ecology. In *Avian Ecology in an Urbanizing World*, edited by J. M. Marzluff, R. Bowman, R. McGowan and R. Donnelly. New York: Kluwer.

Alberti, M., and E. Botsford. 2000. *Behavior of land use and land cover metrics along an urban-rural gradient*. Urban Ecology Research Laboratory Working Paper, Department of Urban Design and Planning, University of Washington.

Alberti, M., and P.A. Waddell. 2000. An integrated urban development and ecological simulation model. *Integrated Assessment*. 1:215-227.

Allen, P.M. 1997. *Cities and Regions as Self-Organizing Systems: Models of Complexity*. Amsterdam: Gordon and Breach Science Publishers.

Allen, P.M., and M. Sanglier. 1979. A dynamic model of growth in a central place system. *Geographical Analysis*. 11 (3):256-272.

Alonso, W. 1960. A theory of the urban land market. *Papers and Proceedings of the Regional Science Association*. 6:149-158.

Alperovich, G., and J. Deutsch. 1992. Population density gradients and urbanisation measurement. *Urban Studies*. 29 (8):1323-1328.

American Farmland Trust. 1995. *Alternatives for Future Urban Growth in California's Central Valley: The Bottom Line for Agriculture and Taxpayers*. Washington, D.C.: American Farmland Trust.

Amson, J.C. 1972. Equilibrium models of cities 1. An axiomatic theory. *Environment and Planning*. 4:429-444.

———. 1973. Equilibrium models of cities 2. Single-species cities. *Environment and Planning*. 5:295-338.

Anderson, J. 2002. Providing a broad spectrum of agents in spatially explicit simulation models: the Gensim approach. In *Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulating Social and Ecological Processes*, edited by H. R. Gimblett. Oxford: Oxford University press 21-55.

Arai, T., and T. Akiyama. 2004. Empirical analysis of the land use transition potential in a CA based land use model: application to the Tokyo Metropolitan Region. *Computers, Environment and Urban Systems*. 28 (1/2):forthcoming.

Audirac, I., A.H. Shermyen, and M.T. Smith. 1990. Ideal urban form and visions of the good life: Florida's growth management dilemma. *Journal of the American Planning Association*. 56 (4):470-482.

Axtell, R. 2000. *Why agents? On the varied motivations for agent computing in the social sciences*. Center on Social and Economic Dynamics Working Paper 17. Washington, D.C.: The Brookings Institute.

Bäck, T., H. Dörnemann, U. Hammel, and P. Frankhauser. 1996. Modeling urban growth by cellular automata. In *Parallel Problem Solving from Nature: PPSN*

IV, edited by H.-M. Voigt, W. Ebeling, I. Rechenberg and H.-P. Schwefel. Berlin: Springer 636-645.

Bae, C.-H.C., and H.W. Richardson. 1994. *Automobiles, the environment and metropolitan spatial structure.* Working Paper Cambridge, MA: Lincoln Institute of Land Policy.

Balchin, P., and J.L. Kieve. 1977. *Urban Land Economics.* London: Macmillan.

Banks, I.M. 1996. *Excession.* London: Orbit.

Barnett, J. 1995. *The Fractured Metropolis: Improving the New City, Restoring the Old City, Reshaping the Region.* New York: IconEditions.

Barrett, C.L., R.J. Beckman, K.P. Berkbigler, K.R. Bisset, B.W. Bush, K. Campbell, S. Eubank, K.M. Henson, J.M. Hurford, D.A. Kubicek, M.V. Marathe, P.R. Romero, J.P. Smith, L.L. Smith, P.L. Speckman, P.E. Stretz, G.L. Thayer, E.V. Eeckhout, and M.D. Williams. 2002. *Chapter 2: Population Synthesizer. TRANSIMS: Transportation Analysis Simulation System. Version: TRANSIMS -3.0. Volume Three: Modules.* Report LA-UR-00-1725. Los Alamos, NM: Los Alamos National Laboratories.

Barrett, C.L., R.J. Beckman, K.P. Berkbigler, K.R. Bisset, B.W. Bush, K. Campbell, S. Eubank, K.M. Henson, J.M. Hurford, D.A. Kubicek, M.V. Marathe, P.R. Romero, J.P. Smith, L.L. Smith, P.E. Stretz, G.L. Thayer, E. Van Eeckhout, and M.D. Williams. 2001. *TRansportation ANalysis SIMulation System (TRANSIMS). Portland Study Reports.* Portland Study Reports LA-UR-01-5711, 5712, 5713, 5714, 5715. Los Alamos: Los Alamos National Laboratory.

Barrett, C.L., R.J. Beckman, K.P. Berkbigler, K.R. Bisset, B.W. Bush, S. Eubank, J.M. Hurford, G. Konjevod, D.A. Kubicek, M.V. Marathe, J.D. Morgeson, M. Rickert, P.R. Romero, L.L. Smith, M.P. Speckman, P.L. Speckman, P.E. Stretz, G.L. Thayer, and M.D. Williams. 1999. *TRANSIMS (TRansportation ANalysis SIMulation System). Volume 0: Overview.* Los Alamos National Laboratory Report LA-UR-99-1658. Los Alamos: Los Alamos National Laboratory.

Batty, M. 1997. Editorial: Urban systems as cellular automata. *Environment and Planning B.* 24:159-164.

———. 1998. Urban evolution on the desktop: simulation with the use of extended cellular automata. *Environment and Planning A.* 30:1943-1967.

———. 2001a. Agent-based pedestrian modeling. *Environment and Planning B.* 28 (3):321-326.

———. 2001b. Polynucleated urban landscapes. *Urban Studies.* 38 (4):635-655.

Batty, M., H. Couclelis, and M. Eichen. 1997. Special issue: urban systems as cellular automata. *Environment and Planning B.* 24 (2).

Batty, M., and S.K. Kwang. 1992. Form follows function: reformulating urban population density functions. *Urban Studies.* 29 (7):1043-1070.

Batty, M., and P. Longley. 1994. *Fractal Cities.* London: Academic Press.

Batty, M., P.A. Longley, and A.S. Fotheringham. 1989. Urban growth and form: scaling, fractal geometry, and diffusion-limited-aggregation. *Environment and Planning A.* 21:1447-1472.

Batty, M., Y. Xie, and Z. Sun. 1999a. *The dynamics of urban sprawl.* CASA Working Paper 15. London: University College London, Centre for Advanced Spatial Analysis (CASA).

———. 1999b. Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban Systems.* 23 (3):205-233.

Baylor University. 2002. GRASS GIS (Geographic Resources Analysis Support System) 5.0. (Software). Waco, TX: Center for Applied Geographic and Spatial Research, Baylor University. Available at http://www.geog.uni-hannover.de/grass/.

Bell, M., C. Dean, and M. Blake. 1999. A model for forecasting the location of fringe urbanisation, with GIS and 3D visualisation. Paper read at Sixth International Conference on Computers in Urban Planning and Urban Management (CUPUM), September, 1999, IUAV, Venice.

Belluck, P. 1999. Board for Kansas deletes evolution from curriculum. *New York Times*, August 12.

Benenson, I. 1998. Multi-agent simulations of residential dynamics in the city. *Computers, Environment and Urban Systems.* 22 (1):25-42.

———. 1999. Modelling population dynamics in the city: from a regional to a multi-agent approach. *Discrete Dynamics in Nature and Society.* 3:149-170.

Benenson, I., and I. Omer. 2000a. Multi-scale approach to measuring residential segregation and the case of Yaffo, Tel-Aviv. In *Proceedings of the Fifth Annual Conference on GeoComputation*, edited by R. Abrahart. Manchester: GeoComputation CD-ROM.

———. 2000b. Studies of unique GIS of Israeli population census: high resolution urban patterns and individual residential segregation. In *Proceedings of the 3rd AGILE Conference on GIS.* Helsinki/Espoo 94-97.

Benenson, I., I. Omer, and E. Hatna. 2002. Entity-based modeling of urban residential dynamics: the case of Yaffo, Tel Aviv. *Environment and Planning B: Planning and Design.* 29:491- 512.

Benenson, I., and P.M. Torrens. 2003. Geographic Automata Systems: a new paradigm for integrating GIS and geographic simulation. Paper read at Association Geographic Information Laboratories Europe (AGILE), April, 2003, Lyons.

———. 2004a. Geosimulation: object-based modeling of urban phenomena. *Computers, Environment and Urban Systems.* 28 (1/2): forthcoming.

———. 2004b. Special Issue: Geosimulation: object-based modeling of urban phenomena. *Computers, Environment and Urban Systems.* 28 (1/2): forthcoming.

Benenson, Y., T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh, and E. Shapiro. 2001. Programmable and autonomous computing machine made of biomolecules. *Nature.* 414:430-434.

Benfield, F.K., M.D. Raimi, and D.D.T. Chen. 1999. *Once there were Greenfields: How Urban Sprawl is Undermining America's Environment, Economy, and*

Social Fabric. New York and Washington, D.C.: Natural Resources Defense Council.

Benguigi, L., D. Czamanski, and M. Marinov. 2001. City growth as a leap-frogging process: an application to the Tel-Aviv metropolis. *Urban Studies.* 38 (10):1819-1839.

Bleicher, N. 1892. *Statistiche Beschreibung der Stadt Franfurt-am-Main und ihrer Bevölkerung.* Frankfurt-am-Main

Blodgett, J. 1998. Geographic Correspondence Engine 3.0. New York: Center for International Earth Science Information Network (CIESIN), Columbia University. Available at http://plue.sedac.ciesin.org/geocorr/.

Bonabeau, E., M. Dorigo, and G. Theraulaz. 1999. *Swarm Intelligence: From Natural to Artificial Systems, Santa Fe Institue Studies in the Sciences of Complexity.* New York: Oxford University Press.

Brail, R.K., and R.E. Klosterman. 2001. *Planning Support Systems in Practice: Integrating Geographic Information Systems, Models, and Visualization Tools.* Redlands, CA and New Brunswick, NJ: ESRI Press and Center for Urban Policy Research Press.

Bramley, G., W. Bartlett, and C. Lambert. 1995. *Planning, the Market and Private Housebuilding, The Natural and Built Environment Series 4.* London: UCL Press.

Brookings Institution. 2001. Ascape 1.9. (Software). Washington, D.C.: Center on Social and Economic Dynamics. Available at http://www.brook.edu/es/dynamics/models/ascape.

Brown, D.G., S.E. Page, R. Riolo, and W. Rand. 2002. Modeling the effects of greenbelts on the urban-rural fringe. Paper read at iEMSs: Integrated Assessment and Decision Support, June 24-27, 2002, Lugano, Switzerland.

Brown, L.A., and E.A. Moore. 1970. The intra-urban migration process: a perspective. *Geografiska Annaler.* 52B:1-13.

Bush, B.W. 2001. *Portland Synthetic Population.* Los Alamos National Laboratory Report LA-UR-00-5972. Los Alamos: Los Alamos National Laboratory.

Bussière, R. 1968. *The spatial distribution of urban populations.* The Hague and Paris: International Federation for Housing and Planning and Le Centre de Recherche d'Urbanisme.

Calthorpe, P., W. Fulton, and R. Fishman. 2001. *The Regional City: Planning for the End of Sprawl.* Washington, D.C.: Island Press.

Camara, A.S., and J. Raper. 1999. Defining spatial agents. In *Spatial Multimedia and Virtual Reality.* London: Taylor and Francis 111-129.

Candau, J.T., S. Rasmussen, and K.C. Clarke. 2000. A coupled cellular automata model for land use/land cover change dynamics. Paper read at 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs, September 2-8, 2000, Banff, Alberta, Canada.

Carter, H. 1981. *The Study of Urban Geography.* Third Edition ed. Victoria, Australia: Edward Arnold.

Casti, J.L. 1997. *Would-be Worlds: How Simulation is Changing the Frontiers of Science*. New York: John Wiley & Sons.

Center for Connected Learning and Computer-Based Modeling. 2001. StarLogoT 2001. Evanston, IL: Northwestern University. Available at http://ccl.sesp.northwestern.edu/cm/starlogoT/.

———. 2003a. HubNet 1.2. (Software). Evanston, IL: Northwestern University. Available at http://ccl.sesp.northwestern.edu/netlogo/.

———. 2003b. NetLogo 1.2. (Software). Evanston, IL: Northwestern University. Available at http://ccl.sesp.northwestern.edu/netlogo/.

Centre for Computational Geography. 2003. GeoTools 2.0. (Software). Leeds: University of Leeds, School of Geography. Available at http://geotools.sourceforge.net/.

Cervero, R. 1998. *Transit Metropolis: A Global Inquiry.* Washington D.C.: Island Press.

Cetin, N., K. Nagel, B. Raney, and A. Voellmy. 2001. Large-scale multi-agent transportation systems. *Computational Physics Communications.* 147 (1-2):559-564. Available from http://www.inf.ethz.ch/personal/nagel/papers/aachen01_CmpPhys/html/.

Chapin, F.S., and S.F. Weiss. 1962. *Factors influencing land development. An Urban Studies Research Monograph.* Chapel Hill: Center for Urban and Regional Studies, Institute for Research in Social Science, University of North Carolina.

———. 1965. *Some input refinements for a residential model. An Urban Studies Research Monograph.* Chapel Hill, North Carolina: Center for Urban and Regional Studies, Institute for Research in Social Science, University of North Carolina.

———. 1968. A Probabilistic Model for Residential Growth. *Transportation Research.* 2:375-390.

Clark, C. 1951. Urban population densities. *Journal of the Royal Statistical Society, Series A.* 114:490-496.

Clark, W.A.V. 1970. Measurement and explanation in intra-urban mobility. *Tijdschrift voor Economische en Sociale Geografie.* 61:49-57.

———. 1982a. Recent research on migration and mobility: a review and interpretation. *Progress and Planning.* 18:1-56.

———. 1993. Search and choice in urban housing markets. In *Behavior and Environment: Psychological and Geographical Approaches*, edited by T. Gärling and R. G. Golledge. Amsterdam: Elsevier/North Holland 298-316.

———, ed. 1982b. *Modelling Housing Market Search.* London: Croom Helm.

Clark, W.A.V., and M. Cadwaller. 1973. Locational stress and residential mobility. *Environment and Behavior.* 5:29-41.

Clark, W.A.V., and R. Flowerdew. 1982. A review of search models and their application to search in the housing market. In *Modelling Housing Market Search*, edited by W. A. V. Clark. London: Croom Helm 4-29.

Clarke, G.P., ed. 1996. *Microsimulation for Urban and Regional Policy Analysis, European Research in Regional Science 6*. London: Pion.

Clarke, K. *Land transition modeling with deltatrons* [Internet]. Department of Geography, University of California, Santa Barbara 1997 [cited. Available from http://www.geog.ucsb.edu/~kclarke/Papers/deltatron.html.

Clarke, K.C., and L. Gaydos. 1998. Loose coupling a cellular automaton model and GIS: long-term growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science.* 12 (7):699-714.

Clarke, K.C., S. Hoppen, and L. Gaydos. 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B.* 24:247-261.

Clawson, M. 1962. Urban sprawl and speculation in suburban land. *Land and Economics.* 38 (2):99-111.

Couclelis, H. 1985. Cellular worlds: A framework for modeling micro-macro dynamics. *Environment and Planning A.* 17:585-596.

———. 1997. From cellular automata to urban models: New principles for model development and implementation. *Environment and Planning B.* 24:165-174.

Coupland, D. 1995. *Microserfs*. New York: Reagan Books.

Cronin, F.J. 1982. Racial differences in the search for housing. In *Modelling Housing Market Search*, edited by W. A. V. Clark. London: Croom Helm 81-106.

De la Barra, T. 1989. *Integrated Land Use and Transport Modelling: Decision Chains and Hierarchies*. Cambridge: Cambridge University Press.

DiPasquale, D., and W. Wheaton. 1996. *Urban Economics and Real Estate Markets*. Englewood Cliffs, N.J.: Prentice Hall.

Duany, A., E. Plater-Zyberk, and J. Speck. 2000. *Suburban Nation: The Rise of Sprawl and the Decline of the American Dream*. New York: North Point Press.

Duany, A., J. Speck, and E. Plater-Zyberk. 2001. *Smart Growth: New Urbanism in American Communities*. New York: McGraw-Hill.

Engelen, G., R. White, and I. Uljee. 2002. *The MURBANDY and MOLAND models for Dublin.* Maastricht: Research Institute for Knowledge Systems (RIKS) BV.

Engelen, G., R. White, I. Uljee, and P. Drazan. 1995. Using cellular automata for integrated modelling of socio-environmental systems. *Environmental Monitoring and Assessment.* 30:203-214.

Environmental Protection Agency. 1993. *Costs of providing government services to alternative residential plans.* Report prepared for the Chesapeake Bay Program's Subcommittee on Population and Growth

———. 2000. *National Air Pollutant Emission Trends: 1900-1998.* EPA-454/R-00-002. Research Triangle Park, NC: U.S. Environmental Protection Agency.

Epstein, J.M. 1999. Agent-based computational models and generative social science. *Complexity.* 4 (5):41-60.

Epstein, J.M., and R. Axtell. 1996. *Growing Artificial Societies from the Bottom Up.* Washington D.C.: Brookings Institution.

Ewing, R. 1994. Causes, characteristics, and effects of sprawl: a literature review. *Environmental and Urban Issues.* 21 (2):1-15.

———. 1997. Is Los Angeles-style sprawl desirable? *Journal of the American Planning Association.* 63 (1):107-126.

Faith, J. 1998. Why Gliders Don't Exist: Anti-Reductionism and Emergence. In *Artificial Life VI: Proceedings of the Sixth International Conference on Artificial Life*, edited by C. Adami. Cambridge, MA: MIT Press.

Fannie Mae Foundation. 1997. *National Housing Survey.* Washington, D.C.: Fannie Mae Foundation.

Farley, R., and W.H. Frey. 1994. Changes in the segregation of whites from blacks in the 1980s: small steps toward a more integrated society. *American Sociological Review.* 59 (23-45).

Farley, R., H. Schuman, S. Binachi, D. Colasanto, and S. Hatchett. 1978. Chocolate city, vanilla suburbs: will the trend toward racially separate communities continue? *Social Science Research.* 7:319-344.

Ferber, J. 1999. *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence.* Harlow (UK): Addison-Wesley.

FHWA. 1995. *Nationwide Personal Travel Survey.* Washington, D.C.: Department of Transportation Federal Highway Administration.

Forrester, J. 1969. *Urban Dynamics.* Cambridge, MA: The MIT Press.

Fotheringham, A.S., C. Brunsdon, and M. Charlton. 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis.* London: Sage.

Fotheringham, A.S., and M.E. O'Kelly. 1989. *Spatial Interaction Models: Formulations and Applications, Studies in Operational Regional Science.* Dordrecht: Kluwer Academic Publishers.

Frank, J.E. 1989. *The Costs of Alternative Development Patterns: A Review of the Literature.* Washington, D.C.: Urban Land Institute.

Fujita, M. , Krugman, P. and A.J. Venables. 2001. *The Spatial Economy: Cities, Regions, and International Trade.* Cambridge, MA: The MIT Press.

Franklin, S., and A. Graesser. 1997. Is it an agent, or just a program?: A taxonomy for autonomous agents. In *Intelligent Agents III: Proceedings of the Third International Workshop on Agent Theories, Architectures and Languages, (ATAL'96), Lecture Notes in AI*, edited by J. P. Mller, M. J. Wooldridge and N. R. Jennings: Springer-Verlag.

Galster, G. 1991. Black suburbanization: has it changed the relative location of the races? *Urban Affairs Quarterly.* 26 (621-628).

Garreau, J. 1992. *Edge City: Life on the New Frontier.* New York: Anchor Books/Doubleday.

Geertman, S., and J. Stillwell, eds. 2002. *Planning Support Systems in Practice.* London: Springer-Verlag.

Gibson, W. 1984. *Neuromancer.* New York: Ace Books.

———. 1993. *Virtual Light.* London: Penguin Books.

———. 1996. *Idoru.* London: Penguin Books.

———. 1999. *All Tomorrow's Parties.* London: Viking.

Gimblett, H.R., ed. 2002. *Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulating Social and Ecological Processes, Santa Fe Institute Studies in the Sciences of Complexity.* Oxford: Oxford University Press.

Giuliano, G. 1989. Research policy and review 27. New directions for understanding transportation and land use. *Environment and Planning A.* 21:145-159.

Gleick, J. 2000. *Faster.* New York: Abacus.

Goldstein, N.C., J.T. Candau, and K.C. Clarke. 2004. Approaches to simulating the "March of Bricks And Mortar". *Computers, Environment and Urban Systems.* 28 (1/2):forthcoming.

Golledge, R., and R.J. Stimson. 1997. *Spatial Behavior: A Geographic Perspective.* New York: The Guilford Press.

Gordon, P., and H.W. Richardson. 1997a. Are compact cities a desirable planning goal? *Journal of the American Planning Association.* 63 (1):95-106.

———. 1997b. Where's the sprawl? *Journal of the American Planning Association.* 63 (2):275-278.

Gottman, J., and R.A. Harper. 1967. *Metropolis on the Move: Geographers Look at Urban Sprawl.* New York: John Wiley & Sons.

Gottmann, J. 1967a. *Megalopolis: The Urbanized Northeastern Seaboard of the United States.* Cambridge, MA: MIT Press.

———. 1967b. Urban sprawl and its ramifications. In *Metropolis on the Move: Geographers Look at Urban Sprawl,* edited by J. Gottman and R. A. Harper. New York: John Wiley & Sons 3-21.

Government of Ireland. 1995. *Dublin Transportation Initiative Technical Volume I: Transport Modelling in Phase 2.* Dublin: Stationary Office.

Guy, S., S. Martin, and T. Moss. 2001. *Urban Infrastructure in Transition: Networks, Buildings, Plans.* London: Earthscan Publications.

Haklay, M., D. O'Sullivan, M. Thurstain-Goodwin, and T. Schelhorn. 2001. "So go downtown": simulating pedestrian movement in town centres. *Environment and Planning B.* 28 (3):343-359.

Hall, P. 1983. Decentralization without end? In *The Expanding City: Essays in Honor of Professor Jean Gottmann.* London: Academic Press.

Handy, S.L., and D.A. Niemeier. 1997. Measuring accessibility: an exploration of issues and alternatives. *Environment and Planning A.* 29:1175-1194.

Harvey, R.O., and W.A.V. Clark. 1965. The nature and econmics of sprawl. *Land Economics.* 61 (1):1-9.

Helbing, D., F. Illés, and T. Vicsek. 2000. Simulating dynamical features of escape panic. *Nature.* 407 (September 28):487-490.

Helbing, D., P. Molnár, I. Farkas, and K. Bolay. 2001. Self-organizing pedestrian movement. *Environment and Planning B.* 28:361-383.

Herold, M. 2002. Remote sensing based analysis of urban dynamics in the Santa Barbara Region using the SLEUTH urban growth model and spatial metrics. Paper read at Third Symposium on Remote Sensing of Urban Areas, June, 2002, Istanbul, Turkey.

Herold, M., and K.C. Clarke. 2002. The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. *Environment and Planning A.* 34:1443-1458.

Hodge, D. 1986. Social impacts of urban transportation decisions: equity issues. In *The Geography of Urban Transportation*, edited by S. Hanson. New York: Guilford Press 301-327.

Holland, J.H. 1995. *Hidden Order: How Adaptation Builds Complexity.* Reading, MA: Addison-Wesley.

———. 1998. *Emergence: From Chaos to Order.* Reading, MA: Perseus Books.

Hoogendorn, S.P., and P.H.L. Bovy. 2002. Normative pedestrian behaviour theory and modelling. In *Transportation and Traffic Theory in the 21st Century*, edited by M. A. P. Taylor. Oxford: Pergamon 219-245.

Horgan, J. 1995. From complexity to perplexity: Can science achieve a unified theory of complexity systems? Even at the Santa Fe Institute, some researchers have their doubts. *Scientific American.* Available from http://www.sciam.com/explorations/0695trends.html.

Horstmann, C.S., and G. Cornell. 2002. *Core Java 2. Volume II: Advanced Features.* Palo Alto: Sun Microsystems Press & Prentice Hall.

Hortsmann, C.S., and G. Cornell. 2001. *Core Java 2, Volume 1: Fundamentals.* Saddle River, NJ: Sun Microsystems Press & Prentice Hall.

HUD. 1999. *The State of the Cities.* Washington, D.C.: U.S. Departmet of Housing and Urban Development.

Huff, J.O. 1982. Spatial aspects of residential search. In *Modelling Housing Market Search*, edited by W. A. V. Clark. London: Croom Helm 106-134.

———. 1986. Geographic regularities in residential search behavior. *Annals of Association of American geographers.* 76:208-227.

Huff, J.O., and W.A.V. Clark. 1978. Cumulative stress and cumulative inertia: a behavioral model of the decision to move. *Environment and Planning A.* 10:1101-1119.

Hughes, M.A., and J.E. Sternberg. 1992. *The New Metropolitan Reality: Where the Rubber Meets the Road in Antipoverty Policy.* Washington, D.C.: The Urban Institute.

Isard, W. 1998. Interregional and spatial microsimulation. In *Methods of Interregional and Regional Analysis*, edited by W. Isard, I. J. Azis, M. P. Drennen, R. E. Miller, S. Saltzman and E. Thorbecke. Aldershot, UK: Ashgate.

Isard, W., I.J. Azis, M.P. Drennen, R.E. Miller, S. Saltzmann, and E. Thorbecke. 1998. *Methods of Interregional and Regional Analysis.* Aldershot, UK: Ashgate.

Jacobs, J. 1961. *The Death and Life of Great American Cities.* New York: Random House.

James Duncan & Associates, G.A. Van Horn, B. Ivey, Harris, and Walls, Inc.,, and I. Wade-Trim. 1989. *The Search for Efficient Urban Growth Patterns: A Study of the Fiscal Impacts of Development in Florida.* Report presented to the Governer's Task Force on Urban Growth Patterns and the Florida Department of Community Affairs

Katz, P. 1993. *The New Urbanism : Toward an Architecture of Community.* New York: McGraw-Hill.

Kendig, H. 1984. Housing careers, life cycle, and residential mobility: implications for the housing market. *Urban Studies.* 4:271-283.

Kerridge, J., J. Hine, and M. Wigan. 2001. Agent-based modelling of pedestrian movements: the questions that need to be asked and answered. *Environment and Planning B.* 28 (3):327-341.

Klosterman, R.E. 1999. The What if? Collaborative Planning Support System. *Environment and Planning B: Planning & Design.* 26:393-408.

———. 2001. The What If? Planning Support System. In *Planning Support Systems: Integrating Geographic Information Systems, Models, and Visualization Tools*, edited by R. K. Brail and R. E. Klosterman. Redlands, CA: ESRI Press 263-284.

Knox, P. 1989. The vulnerable, the disadvantaged, and the victimized: who they are and where they live. In *Social Problems and the City*, edited by D. T. Herbert and D. M. Smith. Oxford: Oxford University Press.

Kohler, T.A. 2000. Putting social sciences together again: an introduction to the volume. In *Dynamics in Human and Primate Societies*, edited by T. A. Kohler and G. Gumerman. New York: Oxford University Press 1-18.

Krugman, P. 1996. *The Self-Organizing Economy.* Malden, MA: Blackwell.

Krummel, J.R., R.H. Gardner, G. Sugihara, and R.V. O'Neill. 1981. Landscape patterns in a disturbed environment. *Oikos.* 48:321-324.

Kunstler, J.H. 1993. *The Geography of Nowhere: The Rise and Decline of America's Man-Made Landscape.* New Yorl: Simon and Schuster.

Kurzweil, R. 1990. *The Age of Intelligent Machines.* Cambridge, Mass.: MIT Press.

———. 1999. *The Age of Spiritual Machines: How we will Live, Work and Think in the New Age of Intelligent Machines.* London: Phoenix.

Landis, J. 1994. The California Urban Futures Model: a new generation of metropolitan simulation models. *Environment and Planning B: Planning & Design.* 21:399-420.

———. 1995. Imagining land use futures: applying the California Urban Futures Model. *Journal of the American Planning Association.* 61 (4):438-457.

———. 2001. CUF, CUF II, and CURBA: a family of spatially explicit urban growth and land-use policy simulation models. In *Planning Support Systems: Integrating Geogrphic Information Systems, Models, and Visualization Tools*, edited by R. K. Brail and R. E. Klosterman. Redlands, CA: ESRI Press 157-200.

Landis, J., and M. Zhang. 1998a. The Second Generation of the California Urban Futures Model. Part 2: specification and calibration results of the land-use change submodel. *Environment and Planning B: Planning & Design.* 25 (6):795-824.

———. 1998b. The Second Generation of the California Urban Futures Model: Part 1: model logic and theory. *Environment and Planning B: Planning & Design.* 25 (5):657-666.

Langton, C.G. 1992. Life at the Edge of Chaos. In *Artificial Life II*, edited by C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen. Redwood City, CA: Addison-Wesley 41-93.

Ledermann, R.C. 1967. The city as a place to live in. In *Metropolis on the Move: Geographers Look at Urban Sprawl*, edited by J. Gottmann and R. A. Harper. New York: John Wiley & Sons.

Leonard, A. 1997. *Bots: The Origin of a New Species.* San Francisco: Hardwired.

Lessinger, J. 1962. The cause for scatteration: some reflections on the National Capitol Region *Plan for the Year 2000. Journal of the American Institute of Planners.* 28 (3):159-170.

Levy, S. 1992. *Artificial Life: The Quest for a New Creation.* Second ed. London: Penguin Books.

Li, H., and J.F. Reynolds. 1994. A simulation experiment to quantify spatial heterogeneity in categorical maps. *Ecology.* 75:2446-2455.

Li, X., and A.G.-O. Yeh. 2000. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science.* 14 (2):131-152.

———. 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science.* 16 (4):323-343.

Longley, P.A., and T.V. Mesev. 2001. Measuring urban morphology using remotely-sensed imagery. In *Remote Sensing and Urban Analysis*, edited by J.-P. Donnay, M. Barnsley and P. A. Longley. London: Taylor and Francis.

Louviere, J.J., D.A. Hensher, and J.D. Swatt. 2000. *Stated Choice Methods: Analysis and Application.* Cambridge: Cambridge University Press.

Lowry, I. 1964. *A Model of Metropolis.* Report RM-4035-RC. Santa Monica: Rand Corporation.

Mayer, H. 1967. The pull of land and space. In *Metropolis on the Move: Geographers Look at Urban Sprawl*, edited by J. Gottmann and R. A. Harper. New York: John Wiley & Sons.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press 105-142.

McGarigal, K., and B. Marks. 1995. *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure.* General Technical Report PNW-GTR-351. USDA Forest Service.

McGarigal, K., B. Marks, C. Holmes, and E. Eve. 2003. FRAGSTATS 3.3 build 4. (Software). Corvallis, OR: Oregon State University, Forest Science Department.

Meyer, J.-A., A. Berthoz, D. Floreano, H.L. Roitblat, and S.W. Wilson, eds. 2000. *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Meyer, J.-A., and A. Guillot. 1994. From SAB90 to SAB94: four years of animat research. In *From Animals to Animats 3. Proceedings of the Third International Conference on Simulation of Adaptive Behavior,* edited by D. Cliff, P. Husbands, J.-A. Meyer and S. Wilson. Cambridge, MA: The MIT Press 2-11.

Microsoft Corporation, and Digital Equipment Corporation. 1995. The Component Object Model Specification 0.9. Redmond, WA: Microsoft Corporation and Digital Equipment Corporation. Available at http://www.microsoft.com/com/resources/comdocs.asp.

Mills, E.S. 1972. *Studies in the Structure of the Urban Economy.* Baltimore: Johns Hopkins Press for Resources for the Future.

Mills, E.S., and J.P. Tan. 1980. A comparison of urban population density functions in developed and developing countries. *Urban Studies.* 17:313-321.

Mitchell, M. 1998. *An Introduction to Genetic Algorithms, Complex Adaptive Systems.* Cambridge, MA: MIT Press.

Moody, G. 2001. *Rebel Code: Linux and the Open Source Revolution.* London: Allen Lane: The Penguin Press.

Morrill, R. 1991. Myths about metropolis. In *Our Changing Cities*, edited by J. F. Hart. Baltimore, MD: Johns Hopkins University Press.

Muth, R. 1969. *Cities and Housing.* Chicago: University of Chicago Press.

Nagel, K., S. Rasmussen, and C.L. Barrett. 1996. *Network traffic as self-organized critical phenomena.* TSA-DO/SA MS-M997 and CNLS MS-B258. Los Alamos, NM: Los Alamos National Laboratory.

Nagel, K., and M. Rickert. 2001. Parallel implementation of the TRANSIMS micro-simulation. *Parallel Computing.* 27 (12):1611-1639.

Nagel, K., and M. Schrekenberg. 1995. *Traffic jams in stochastic cellular automata.* Los Alamos National Laboratory Report 95ATS089. Los Alamos, NM: Los Alamos National Laboratory.

Nakajima, T. 1977. Application de la théorie de l'automate à la simulation de l'évolution de l'espace urbain. In *Congrès Sur La Méthodologie De L'Aménagement Et Du Dévelopment.* Montreal: Association Canadienne-Française Pour L'Avancement Des Sciences et Comité De Coordination Des

Centres De Recherches En Aménagement, Développement Et Planification (CRADEP) 154-160.

Noon, J. 1993. *Vurt.* New York: Crown Publishers, Inc.

Noth, M., A. Borning, and P.A. Waddell. 2003. An extensible, modular architecture for simulating urban development, transportation, and environmental impacts. *Computers, Environment and Urban Systems.* 27:181-203.

O'Flaherty, B. 1996. *Making Room: The Economics of Homelessness.* Cambridge, MA: Harvard University Press.

Openshaw, S. 1983. *The Modifiable Areal Unit Problem, CATMOG 38.* Norwich: GeoBooks.

Orcutt, G. 1957. A new type of socio-economic systems. *The Review of Economics and Statistics.* 58:773-797.

Orcutt, G., S. Caldwell, and R. Wertheimer. 1976. *Policy Exploration through Microanalytic Simulation.* Wahington, D.C: The Urban Institute.

Orcutt, G., M. Greenberger, J. Korbel, and A. Rivlin. 1961. *Microanalysis of Socioeconomic Systems: a Simulation Study.* New York: Harper & Row.

Orfield, M. 1997. *Metropolitics: A Regional Agenda for Community and Stability.* Washington, D.C.: Brookings Press.

O'Sullivan, D. 2001. Exploring spatial process dynamics using irregular cellular automaton models. *Geographical Analysis.* 33 (1):1-18.

O'Sullivan, D., and P.M. Torrens. 2000. Cellular models of urban systems. In *Theoretical and Practical Issues on Cellular Automata*, edited by S. Bandini and T. Worsch. London: Springer-Verlag 108-117.

OTA. 1995. *The Technological Reshaping of Metropolitan America.* OTA-ETI-643. Washington, D.C.: U.S. Congress Office of Technology Assessment.

Ottensmann, J.R. 1977. Urban sprawl, land values and the density of development. *Land Economics.* 53 (4):389-400.

Pallman, D. 1999. *Programming Bots, Spiders, and Intelligent Agents in Microsoft Visual C++, Microsoft Programming Series.* Redmond, WA: Microsoft Press.

Palm, R. 1982. Homebuyer response to information content. In *Modelling Housing Market Search*, edited by W. A. V. Clark. London: Croom Helm 187-209.

Parker, D.C., S.M. Manson, M.A. Janssen, M.J. Hoffmann, and P. Deadman. 2003. Multi-Agent System models for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers.* 93 (2):314-337.

Peiser, R. 1989. Density and urban sprawl. *Land Economics.* 65 (3):193-204.

Pendall, R. 1999. Do land-use controls cause sprawl? *Environment and Planning B.* 26:555-571.

Peterson, G.E. 1980. Federal tax policy and the shaping of urban development. In *The Prospective City*, edited by A. P. Solomon. Cambridge, MA: MIT Press 48-49.

Phipps, M. 1989. Dynamic behavior of cellular automata under the constraint of neighborhood coherence. *Geographical Analysis.* 21:197-215.

Pickvance, G.C. 1973. Life-cycle, housing tenure and intra-urban residential mobility: a causal model. *American Sociological Review.* 21:279-297.

Portugali, J. 2000. *Self-Organization and the City.* Berlin: Springer-Verlag.

Portugali, J., I. Benenson, and I. Omer. 1994. Socio-spatial residential dynamics: stability and instability within a self-organized city. *Geographical Analysis.* 26 (4):321-340.

———. 1997. Spatial cognitive dissonance and sociospatial emergence in a self-organizing city. *Environment and Planning B.* 24:263-285.

Poundstone, W. 1985. *The Recursive Universe : Cosmic Complexity and the Limits of Scientific Knowledge*. New York: Morrow.

Preece, J. 1994. *Human-Computer Interaction.* Harlow, England: Addison-Wesley Publishing.

Rand, W., M. Zellner, S.E. Page, R. Riolo, D.G. Brown, and L. Fernandez. 2002. The complex interaction of agents and environments: an example in urban sprawl. Paper read at Agent 2002, Chicago, IL.

Rasband, W. 2003. Image J 1.29x. (Software). Washington D.C.: National Institutes of Health.

Real Estate Research Corporation. 1974. *The Costs of Sprawl: Environmental and Economic Costs of Alternative Residential Patterns at the Urban Fringe.* Washington, D.C.: U.S. Government Printing Office.

Resnick, M. 1994a. Changing the centralized mind. *Technology Review.* (July):32-40.

———. 1994b. Learning about life. *Artificial Life.* 1 (1-2).

———. 1996. Beyond the centralized mindset. *Journal of the Learning Sciences.* 5 (1):1-22.

———. 1997. *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds.* Fifth ed. Cambridge, MA: MIT Press.

Reynolds, C. 1987. Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics.* 21 (4):25-34.

Richardson, H.W. 1977. *The New Urban Economics: and Alternatives.* London: Pion.

Rickert, M., K. Nagel, M. Schreckenberg, and A. Latour. 1996. Two lane traffic simulations using cellular automata. *Physica A.* (231):534-550.

Rossi, P.H. 1955. *Why Families Move: A Study in the Social Psychology of Urban Residential Mobility.* Glencoe, IL: Free Press.

Rucker, R. 1999. *Seek! Selected Nonfiction by Rudy Rucker.* New York: Four Walls Eight Windows.

Russell, S., and P. Norvig. 1995. *Artificial Intelligence: A Modern Approach.* Upper Saddle River, New Jersey: Prentice-Hall.

Sakoda, J.M. 1971. The checkerboard model of social interaction. *Journal of Mathematical Sociology.* 1:119-132.

Sanders, L., D. Pumain, H. Mathian, F. Guérin-Pace, and S. Bura. 1997. SIMPOP: A multiagent system for the study of urbanism. *Environment and Planning B.* 24:287-305.

Schelling, T.C. 1969. Models of segregation. *American Economic Review.* 59 (2):488-493.

———. 1971. Dynamic models of segregation. *Journal of Mathematical Sociology.* 1:143-186.

———. 1974. On the ecology of micro-motives. In *The Corporate Society*, edited by R. Marris. London: Macmillan 19-55.

———. 1978. *Micromotives and Macrobehavior.* New York: WW Norton and Company.

Semboloni, F. 1997. An urban and regional model based on cellular automata. *Environment and Planning B.* 24:589-612.

———. 2000. The dynamic of an urban cellular automata model in a 3-D spatial pattern. Paper read at XXI National Conference Aisre: Regional and Urban Growth in a Global Market, September 20 to 22, Palermo.

Sermons, M.W. 2000. Influence of race on household residential utility. *Geographical Analysis.* 32 (3):225-246.

Shi, W., and M.Y.C. Pang. 2000. Development of Voronoi-based cellular automata--an integrated dynamic model for Geographical Information Systems. *International Journal of Geographical Information Science.* 14 (5):455-474.

Shiode, N., and P.M. Torrens. 2003a. A comparative study on the growth of real and virtual cities. Paper read at Computers in Urban Planning and Urban Management (CUPUM), June, 2003, Sendai, Japan.

———. 2003b. A preliminary analysis on the fractal structure of urban sprawl dynamics. Paper read at The Sixth Sharjah Urban Planning Symposium, June, 2003, Dubai.

Silva, E.A., and K.C. Clarke. 2002. Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems.* 26 (6):525-552.

Simon, H.A. 1956. Rational choice and the structure of the environment. *Psychological Review.* 63:129-138.

Sipper, M. 1997. *Evolution of Parallel Cellular Machines: The Cellular Programming Approach, Lecture Notes in Computer Science 1194.* Berlin: Springer.

Smeed, R.J. 1963. The effect of some kinds of routeing systems on the amount of traffic in the central areas of towns. *Journal of the Institution of Highway Engineers.* 10 (1):5-26.

Soja, E. 1995. Postmodern urbanization: the six restructurings of Los Angeles. In *Postmodern Cities and Spaces*, edited by S. Watson and K. Gibson. Oxford: Blackwell.

Speare, A., S. Goldstein, and W.H. Frey. 1975. *Residential Mobility, Migration and Metropolitan Change.* Cambridge, MA: Ballinger Publishing Company.

Stephenson, N. 1993. *Snow Crash.* New York: Bantam Books.

———. 1995. *The Diamond Age.* London: Penguin Books.

Sutton, P. 1997. Modeling population density with night-time satellite imagery and GIS. *Computers, Environment and Urban Systems.* 21 (3/4):227-244.

Swarm Development Group. 2001. Swarm 2.1.1. (Software). Santa Fe: Swarm Development Group. Available at http://www.swarm.org.

Taylor, C.E. 1992. "Fleshing Out" Artificial Life II. In *Artificial Life II: Proceedings of the Workshop on Artificial Life held February, 1990 in Santa Fe, New Mexico*, edited by C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen. Redwood City, CA: Addison-Wesley 25-38.

Terna, P. 1998. Simulation tools for social scientists: Building agent based models with SWARM. *Journal of Artificial Societies and Social Simulation.* 1 (2). Available from http://www.soc.surrey.ac.uk.JASSS/1/2/4.html.

Thomas, R.M., and R.J. Huggett. 1980. *Modelling in Geography: A Mathematical Approach.* Totowa, N.J.: Barnes & Noble Books.

Tobler, W. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography.* 46 (2):234-240.

———. 1979. Cellular Geography. In *Philosophy in Geography,* edited by S. Gale and G. Ollson. Dordrecht: Kluwer 379-386.

Torrens, P.M. 1998. Everything Must Go! The Geography of Retail Change in America's Northeastern Megalopolis, 1977 to 1992. Master's Thesis, Department of Geography, Indiana University, Bloomington.

———. 2000a. *How cellular models of urban systems work.* CASA Working Paper 28. London: University College London, Centre for Advanced Spatial Analysis.

———. 2000b. *How Land-Use and Transportation Models Work.* CASA Working Paper 20. London: University College London, Centre for Advanced Spatial Analysis.

———. 2001. *New tools for simulating housing choices.* Program on Housing and Urban Policy Conference Paper Series C01-006. Berkeley, CA: University of California Institute of Business and Economic Research and Fisher Center for Real Estate and Urban Economics.

———. 2002. Cellular automata and multi-agent systems as planning support tools. In *Planning Support Systems in Practice*, edited by S. Geertman and J. Stillwell. London: Springer-Verlag 205-222.

———. 2003a. Automata-based models of urban systems. In *Advanced Spatial Analysis*, edited by M. Batty. Redlands, CA: ESRI Press.

———. 2004a. Geosimulation approaches to traffic modeling. In *Transport Geography and Spatial Systems*, edited by P. Stopher, K. Button, K. Haynes and D. Hensher. London: Pergamon, forthcoming.

———. 2004b. Looking forward: remote sensing as dataware for human settlement simulation. In *Remote Sensing of Human Settlements*, edited by M. Ridd. New York: John Wiley and Sons, forthcoming.

Torrens, P.M., and M. Alberti. 2000. *Measuring sprawl.* CASA Working Paper 27. London: University College London, Centre for Advanced Spatial Analysis.

Torrens, P.M., and I. Benenson. 2003. Geographic Automata Systems. Paper read at Special session: Geographical Perspectives on Complexity Theory and Complex Systems 2, 99th Annual Meeting of the Association of American Geographers, March, 2003, New Orleans.

———. 2004. *Geosimulation: Automata-Based Modeling of Urban Phenomena.* London: John Wiley & Sons.

Torrens, P.M., and D. O'Sullivan. 2000a. Cities, cells, and cellular automata: Developing a research agenda for urban geocomputation. Paper read at Fifth Annual Conference on GeoComputation, Greenwich.

———. 2000b. Cities, cells, and cellular automata: Developing a research agenda for urban geocomputation. In *Proceedings of the Fifth Annual Conference on GeoComputation*, edited by R. Abrahart. Manchester: GeoComputation CD-ROM.

———. 2001. Cellular automata and urban simulation: where do we go from here? *Environment and Planning B.* 28 (2):163-168.

Tu, Y., and J. Goldfinch. 1996. A two-stage housing choice forecasting model. *Urban Studies.* 33 (3):517-537.

Turner, A., and A. Penn. 2002. Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment. *Environment and Planning B: Planning and Design.* 29 (4):473 - 490.

Turner, M.G. 1989. Landscape ecology: The effect of pattern on process. *Annual Review of Ecological Systems.* 20:171-197.

U.S. Census Bureau. 1996. *Current Population Reports. Projections of the Number of Households and Families in the United States: 1995 to 2010.* Washington, D.C.: U.S. Department of Commerce, Economics and Statistics Administration.

Ungerer, M.J., and M.F. Goodchild. 2002. Integrating spatial data analysis and GIS: a new implementation using the Component Object Model (COM). *International Journal of Geographical Information Science.* 16 (1):41-53.

University of Chicago. 2003. RePast 2.0. (Software). Chicago: Social Science Research Computing Program. Available at http://repast.sourceforge.net.

Vicsek, T. 2001. A question of scale. *Nature.* 411 (24 May):421.

Waddell, P.A. 2000. A behavioural simulation model for metropolitan policy analysis and planning: residential location and housing market components of UrbanSim. *Environment and Planning B.* 27 (2):247-263.

———. 2001. Between politics and planning: UrbanSim as a decision-support system for metropolitan planning. In *Planning Support Systems: Integrating Geographic Information Systems, Models, and Visualization Tools,* edited by R. K. Brail and R. E. Klosterman. Redlands, CA: ESRI Press 201-228.

———. 2002. UrbanSim: modeling urban development for land use, transportation and environmental planning. *Journal of the American Planning Association.* 68 (3):297-314.

Waddell, P.A., A. Borning, M. Noth, N. Freier, M. Becke, and G. Ulfarsson. 2003. Microsimulation of urban development and location choices: design and implementation of UrbanSim. *Networks and Spatial Economics.* 3 (1):43-67.

Wagner, D.F. 1997. Cellular automata and geographic information systems. *Environment and Planning B.* 24:219-234.

Wagner, P., K. Nagel, and D.E. Wolf. 1997. Realistic multi-lane traffic rules for cellular automata. *Physica A.* 234:687-698.

Wang, F., and Y. Zhou. 1999. Modelling urban population densities in Beijing 1982-90: suburbanisation and its causes. *Urban Studies.* 36 (2):271-287.

Ward, D.P., A.T. Murray, and S.R. Phinn. 2000. A stochastically constrained cellular model of urban growth. *Computers, Environment and Urban Systems.* 24:539-558.

Ward, D.P., S.R. Phinn, and A.T. Murray. 2000. Monitoring growth in rapidly urbanizing areas using remotely sensed data. *The Professional Geographer.* 52 (3).

Webster, C., F. Wu, and S. Zhou. 1998. An object-based urban simulation model for interactive visualisation. Paper read at Third International Conference on GeoComputation, 17-19 September, 1998, University of Bristol.

Webster, C.J. 1995. Urban morphological fingerprints. *Environment and Planning B.* 22:279-297.

Webster, C.J., and F. Wu. 1998. Simulations of urban growth with models of pollution property rights and subcentre formation. In *Graphics, Visualisation and the Social Sciences: Report from Workshop held May 8-9. Advisory Group on Computer Graphics (AGOCG) Technical Report Series No. 33*, edited by A. Mumford. Loughborough: Loughborough University Joint Information Systems Committee (JISC) 113-122.

———. 1999a. Regulation, land use mix and urban performance. Part 1, performance. *Environment and Planning A.* 31 (8):1433-1442.

———. 1999b. Regulation, land use mix and urban performance. Part 2, theory. *Environment and Planning A.* 31 (9):1529-1547.

Wegener, M. 1983. *Description of the Dortmund Region Model.* Working Paper 8. Dortmund: Institut für Raumplanung.

———. 1994. Operational urban models: state of the art. *Journal of the American Planning Association.* 60:17-29.

———. 1996. The potential of microsimulation for urban models. In *Microsimulation for Urban and Regional Policy Analysis*, edited by G. P. Clarke. London: Pion 149-163.

White, R. 1998. Cities and cellular automata. *Discrete Dynamics in Nature and Society.* 2:111-125.

White, R., and G. Engelen. 1993. Cellular automata and fractal urban form. *Environment and Planning A.* 25:1175-1199.

———. 1994. Urban systems dynamics and cellular automata: fractal structures between order and chaos. *Chaos, Solitions, and Fractals.*

———. 1997. Cellular automata as the basis of integrated dynamic regional modelling. *Environment and Planning B.* 24:235-246.

———. 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems.* 24:383-400.

White, R., G. Engelen, and I. Uljee. 1997. The use of constrained cellular automata for high-resolution modelling of urban land use dynamics. *Environment and Planning B.* 24:323-343.

Wilson, A.G. 1970. *Entropy in Urban and Regional Modelling.* London: Pion.

Wolfram, S. 1994. *Cellular Automata and Complexity.* Reading, MA: Addison-Wesley.

———. 2002. *A New Kind of Science.* Champaign, IL: Wolfram Media, Inc.

Wolpert, J. 1965. Behavioral aspects of the decision to migrate. *Papers and Proceedings of the Regional Science Association.* 15:159-169.

———. 1966. Migration as an adjustment to environmental stress. *Journal of Social Issues.* 22:92-102.

Wrigley, N., T. Holt, D. Steel, and M. Tranmer. 1996. Analysing, modelling, and resolving the ecological fallacy. In *Spatial Analysis: Modelling in a GIS Environment*, edited by P. A. Longley and M. Batty. Cambridge: GeoInformation International.

Wu, F. 1998. An experiment on the generic polycentricity of urban growth in a cellular automatic city. *Environment and Planning B.* 25:731-752.

———. 2002. Calibration of stochastic cellular automata: the application to rural-urban land conversions. *International Journal of Geographical Information Science.* 16 (8):795-818.

Wu, F., and C.J. Webster. 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B.* 25:103-126.

———. 2000. Simulating artificial cities in a GIS environment: urban growth under alternative regulation regimes. *International Journal of Geographical Information Science.* 14 (7):625-648.

Wylie, B., G. Cameron, W. Matthew, and D. McArthur. 1993. PARAMICS: Parallel Microscopic Traffic Simulator. Paper read at Second European Connection Machine Users Meeting, 11-14 October, Meudon, Paris.

Xie, Y. 1996. A generalized model for cellular urban dynamics. *Geographical Analysis.* 28 (4):350-373.

Yeh, A.G.-O. 1998. Sustainable land development model for rapid growth areas using GIS. *International Journal of Geographical Information Systems.* 12 (2):169-189.

Yeh, A.G.-O., and X. Li. 2000. Simulation of compact cities based on the integration of cellular automata and GIS. In *Theoretical and Practical Issues on Cellular Automata*, edited by S. Bandini and T. Worsch. London: Springer-Verlag 170-178.

———. 2001. A constrained CA model for the simulation and planning of sustainable urban forms by using GIS. *Environment and Planning B: Planning & Design.* 28 (5):733-753.

———. 2002. A cellular automata model to simulate development density for urban planning. *Environment and Planning B: Planning & Design.* 29 (3):431-450.

Zielinski, K. 1979. Experimental analysis of eleven models of urban population density. *Environment and Planning A.* 11:629-641.